

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

**A Thesis Submitted for the Degree of PhD at the University of Warwick**

<http://go.warwick.ac.uk/wrap/45785>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

## Library Declaration and Deposit Agreement

### 1. STUDENT DETAILS

Please complete the following:

Full name: ..... John Kenneth Pearman .....

University ID number: ..... 0756323 .....

### 2. THESIS DEPOSIT

2.1 I understand that under my registration at the University, I am required to deposit my thesis with the University in BOTH hard copy and in digital format. The digital version should normally be saved as a single pdf file.

2.2 The hard copy will be housed in the University Library. The digital version will be deposited in the University's Institutional Repository (WRAP). Unless otherwise indicated (see 2.3 below) this will be made openly accessible on the Internet and will be supplied to the British Library to be made available online via its Electronic Theses Online Service (EThOS) service.

[At present, theses submitted for a Master's degree by Research (MA, MSc, LLM, MS or MMedSci) are not being deposited in WRAP and not being made available via EthOS. This may change in future.]

2.3 In exceptional circumstances, the Chair of the Board of Graduate Studies may grant permission for an embargo to be placed on public access to the hard copy thesis for a limited period. It is also possible to apply separately for an embargo on the digital version. (Further information is available in the *Guide to Examinations for Higher Degrees by Research*.)

2.4 If you are depositing a thesis for a Master's degree by Research, please complete section (a) below. For all other research degrees, please complete both sections (a) and (b) below:

#### (a) Hard Copy

I hereby deposit a hard copy of my thesis in the University Library to be made publicly available to readers (please delete as appropriate) EITHER immediately OR after an embargo period of ..... months/years as agreed by the Chair of the Board of Graduate Studies.

I agree that my thesis may be photocopied.

YES / ~~NO~~ (Please delete as appropriate)

#### (b) Digital Copy

I hereby deposit a digital copy of my thesis to be held in WRAP and made available via EThOS.

Please choose one of the following options:

EITHER My thesis can be made publicly available online. YES / ~~NO~~ (Please delete as appropriate)

OR My thesis can be made publicly available only after.....[date] (Please give date)

YES / NO (Please delete as appropriate)

OR My full thesis cannot be made publicly available online but I am submitting a separately identified additional, abridged version that can be made available online.

YES / NO (Please delete as appropriate)

OR My thesis cannot be made publicly available online.

YES / NO (Please delete as appropriate)

### 3. GRANTING OF NON-EXCLUSIVE RIGHTS

Whether I deposit my Work personally or through an assistant or other agent, I agree to the following:

Rights granted to the University of Warwick and the British Library and the user of the thesis through this agreement are non-exclusive. I retain all rights in the thesis in its present version or future versions. I agree that the institutional repository administrators and the British Library or their agents may, without changing content, digitise and migrate the thesis to any medium or format for the purpose of future preservation and accessibility.

### 4. DECLARATIONS

(a) I DECLARE THAT:

- I am the author and owner of the copyright in the thesis and/or I have the authority of the authors and owners of the copyright in the thesis to make this agreement. Reproduction of any part of this thesis for teaching or in academic or other forms of publication is subject to the normal limitations on the use of copyrighted materials and to the proper and full acknowledgement of its source.
- The digital version of the thesis I am supplying is the same version as the final, hard-bound copy submitted in completion of my degree, once any minor corrections have been completed.
- I have exercised reasonable care to ensure that the thesis is original, and does not to the best of my knowledge break any UK law or other Intellectual Property Right, or contain any confidential material.
- I understand that, through the medium of the Internet, files will be available to automated agents, and may be searched and copied by, for example, text mining and plagiarism detection software.

(b) IF I HAVE AGREED (in Section 2 above) TO MAKE MY THESIS PUBLICLY AVAILABLE DIGITALLY, I ALSO DECLARE THAT:

- I grant the University of Warwick and the British Library a licence to make available on the Internet the thesis in digitised format through the Institutional Repository and through the British Library via the EThOS service.
- If my thesis does include any substantial subsidiary material owned by third-party copyright holders, I have sought and obtained permission to include it in any version of my thesis available in digital format and that this permission encompasses the rights that I have granted to the University of Warwick and to the British Library.

### 5. LEGAL INFRINGEMENTS

I understand that neither the University of Warwick nor the British Library have any obligation to take legal action on behalf of myself, or other rights holders, in the event of infringement of intellectual property rights, breach of contract or of any other right, in the thesis.

---

*Please sign this agreement and return it to the Graduate School Office when you submit your thesis.*

Student's signature:  Date: 16/01/2012

# **Molecular Ecology and Transcriptomics of Marine Photosynthetic Picoeukaryotes**

John Kenneth Pearman

A thesis submitted in fulfilment of the requirements for the degree of  
Doctor of Philosophy

**University of Warwick, Coventry, United Kingdom**

**Department of Life Sciences**

January 2012



## Table of contents

<b>List of Figures</b>	vii
<b>List of Tables</b>	x
<b>Acknowledgements</b>	xiii
<b>Declaration</b>	xiv
<b>Summary</b>	xv
<b>Abbreviations</b>	xvi
<b>1. Introduction</b>	1
1.1. Evolution of photosynthesis	2
1.2. Global carbon fixation	3
1.3. Primary production in the marine system	6
1.4. Control of primary production	8
1.5. Biogeography of the marine environment	10
1.6. Plastid evolution as a basis for photosynthetic eukaryote diversity	12
1.7. Diversity of photosynthetic picoeukaryotes (PPEs)	15
1.8. Genomics of marine phytoplankton	23
1.9. History of microbial environmental metagenomics	29
1.10. Metagenomic of marine phytoplankton	32
1.11. Metatranscriptomics of marine phytoplankton	35
1.12. Overall aims	36
1.12.1. Specific aims	37
<b>2. Methods</b>	39
2.1. Culturing	40
2.1.1. Growth and maintenance of <i>E. coli</i> strains	40
2.1.2. Isolation of photosynthetic picoeukaryote (PPE) cultures	41
2.1.2.1. Collection of seawater samples for culturing during the AMT18 cruise	41
2.1.2.2. PPE cell sorting using flow cytometry	41
2.1.2.3. Concentration from natural seawater samples	41
2.1.2.4. Growth of PPEs	42
2.1.3. Maintenance of PPE cultures	42
2.1.4. Purification of agar	42
2.1.5. Agar plating	44
2.1.5.1. Preparation of agar	44
2.1.5.2. Culturing of picoeukaryotes using pour plates	44
2.1.5.3. Culturing of picoeukaryotes using solid agar.	44
2.2. Molecular Techniques	45
2.2.1. DNA extraction protocols	45
2.2.1.1. Chemical denaturation for cultures	45

2.2.1.2. Nucleic acid extraction for sorted samples from the SOLAS cruise (2008)	45
2.2.1.3. Nucleic acid extraction for flow cytometry sorted picoeukaryotes cells from the AMT cruise (2008)	46
2.2.1.4. DNA extraction for total phytoplankton filters from the AMT cruise	46
2.2.2 Polymerase Chain Reaction (PCR)	47
2.2.2.1 Conditions for plastid-biased 16S rRNA PCR	47
2.2.2.2 Conditions for the amplification of picoeukaryotic nuclear 18S rDNA	48
2.2.2.3 Conditions for the specific amplification of prymnesiophyte sequences	49
2.2.3 PCR purification	51
2.2.4 Construction of clone libraries	51
2.2.4.1 Construction of clone libraries using DNA extracted from sorted cells from SOLAS D326	51
2.2.4.2 Construction of clone libraries from DNA extracted from AMT18 samples	52
2.2.5 Restriction endonuclease digestion	53
2.2.6 Sequencing of clone library inserts	53
2.3 Transcriptomic techniques	53
2.3.1 Sample collection and culture maintenance	53
2.3.1.1 Ochromonas sp. (CCMP 584) culture conditions	53
2.3.1.2 Clade VIIA prasinophyte (RCC1124) culture conditions	54
2.3.1.3 Collection of environmental RNA samples	54
2.3.1.4 Flow cytometry sorting of environmental transcriptomic samples	54
2.3.2 Total RNA extraction	55
2.3.3 mRNA purification for cultures	56
2.3.4 Enrichment of mRNA from sorted samples	56
2.3.5 cDNA library construction for mRNA from cultures	57
2.3.6 mRNA amplification from sorted samples	58
2.3.7 Sequencing of cDNA using 454	60
2.4 Statistical analysis	60
2.4.1 Species richness estimations	60
2.4.1.1 Non parametric estimators of species richness	60
2.4.1.2 Coverage values of clone libraries	61
2.4.1.3 Species diversity estimation using Margalef's diversity index	61
2.4.1.4 Comparison of species composition with Morisita-Horn	61
2.4.1.5 Hierarchical cluster analysis	62
2.4.1.6 Canonical correspondence analysis (CCA) and Variation Partitioning Analysis (VPA)	62
2.4.2 Phylogenetic analysis	63

2.4.2.1 Sequence assembly	63
2.4.2.2 Chimera check	63
2.4.2.3 Phylogenetic analysis	63
2.4.3 Transcriptomical analysis	64
2.4.3.1 Construction of contigs from raw 454 reads	64
2.4.3.2 Analysis of transcriptomic reads against the NCBI nr database	64
2.4.3.3 Rapid Analysis of Multiple Metagenomes with a Clustering and Annotation Pipeline (RAMMCAP)	64
2.4.3.4 Annotation of ORFs using Cluster of orthologous groups for eukaryotic genomes (KOG)	66
2.4.3.5 Annotation of ORFs using Kyoto Encyclopedia of Genes and Genomes (KEGG)	66
2.4.3.6 Metagenome Analyze (MEGAN)	66
<b>3. Significant CO<sub>2</sub> fixation by small prymnesiophytes in the Northeast Atlantic Ocean – the SOLAS cruise transect</b>	67
3.1. Introduction	68
3.2. Materials and methods	70
3.2.1. Collection of samples from the D326 cruise	70
3.2.2. Extraction of environmental DNA	72
3.2.3. PCR conditions & construction of clone libraries	72
3.2.4. Simpsons Diversity Index	72
3.3. Results	72
3.3.1. Phylogenetic composition of Euk-A PPEs at station 16400 of the D326 cruise.	72
3.3.2. Phylogenetic composition of Euk-B PPEs at station 16400 of the D326 cruise.	73
3.3.3 Diversity statistics	76
3.4. Discussion	77
<b>4. Community structure of flow-sorted PPE populations along an Atlantic Meridional Transect</b>	80
4.1. Introduction	81
4.2. Methods	85
4.2.1. Cruise details	85
4.2.2. Enumeration of PPEs during the Atlantic Meridional Transect	86
4.2.3. Sample collection for bulk community DNA	86
4.2.4. Collection of samples for flow cytometric sorting of picoeukaryotes	87
4.2.5. Flow cytometric sorting of picoeukaryotes for nucleic acid extraction	87
4.2.6. DNA extraction from filters and sorted cells	88
4.2.7. PCR amplification of 16S and 18S rRNA genes	88
4.2.8. Construction of clone libraries	88

4.2.9. Phylogenetic analysis	89
4.2.10. Statistical analysis	89
4.3. Results	90
4.3.1. Ancillary measurements taken during AMT18	90
4.3.2. Taxonomic composition of total PPEs from bulk community filters	96
4.3.3. Taxonomic diversity of Euk-A and Euk-B PPE populations along AMT18	106
4.3.4. Comparison of PPE community structure as assessed using environmental DNA extracted from bulk community filters or from flow cytometrically sorted cells.	118
4.3.5. Species richness of flow cytometrically sorted clone libraries along the AMT18 latitudinal gradient	119
4.3.6. The composition of picoeularyotes using 18S rRNA gene clone libraries from flow cytometry sorted cells	121
4.3.7. A focused investigation of the diversity of Prymnesiophyceae at a single location in the southern gyre along AMT18 using class-specific primers on flow cytometry sorted Euk-A cells	126
4.3.8. Comparison of PPE diversity along AMT18 using the various plastid and nuclear targeted PCR primer sets	127
4.3.9 Canonical correspondence analysis (CCA) of PPE community structure determined from bulk community DNA samples and measured environmental variables along AMT18	127
4.3.10 CCA of PPE community structure determined from flow cytometry sorted cells and measured environmental variables along AMT18	129
4.4 Discussion	132
4.4.1 PPE abundance along AMT18	132
4.4.2 Diversity of PPEs using marine algal plastid biased 16S rRNA PCR primers	133
4.4.3 Diversity of PPEs along AMT18 as evidenced using nuclear 18S rRNA gene primers	137
4.4.4 Comparison of PPE community structure along AMT18 using plastid and nuclear rRNA clone libraries on flow cytometrically sorted cells	140
4.4.5 Latitudinal diversity gradient	143
4.4 Summary	147
<b>5. Approaches to isolate novel PPEs from AMT18</b>	<b>149</b>
5.1. Introduction	150
5.2. Materials and methods	152
5.2.1. Collection of PPEs for culture	152
5.2.2. Preparation of growth media used to culture PPEs	153
5.2.3. Soil extract preparation	154
5.2.4. Colloidal iron preparation	155
5.2.5. Obtaining clonal cultures of PPEs	155
5.2.6. Longer-term maintenance of PPE cultures	155
5.2.7. Molecular characterisation of cultures	155



5.3. Results	156
5.3.1. Isolation of PPEs enriched using soil extract medium	156
5.3.2. Isolation of cultures from seawater concentrates enriched with f/2 medium	159
5.3.3. Isolation of cultures enriched with artificial K medium	160
5.3.4. Isolation of cultures enriched with artificial f/2 medium following flow cytometric sorting of natural PPE communities obtained on AMT18	160
5.3.5. Isolation of clonal cultures from PPE enrichments	163
5.3.6. Isolation of PPEs into culture: a comparison of different media	163
5.4. Discussion	164
5.5. Conclusions and future directions	167

## **6. Transcriptomics of cultured marine PPEs focusing on a clade VII**

<b>prasinophyte and a marine <i>Ochromonas</i> sp. (Chrysophyceae)</b>	173
6.1. Introduction	174
6.2. Materials and Methods	177
6.2.1. Cultures used for transcriptomic analysis	177
6.2.2. Methods for RNA extraction	178
6.3.3. Enrichment of mRNA from the total RNA	178
6.3.4. Construction of cDNA libraries from enriched eukaryotic mRNA	178
6.3.5. Sequencing of cDNA	179
6.3.6. Phylogenetic analysis of cDNA contigs	179
6.3.7. Functional prediction of cDNA reads	179
6.3. Results	180
6.3.1. Comparison between RNA extraction methods	180
6.3.2. Construction of cDNA library	181
6.3.3. Sequencing results	181
6.3.4. Phylogenetic analysis of transcriptomes	183
6.3.5. Classification of ORFs against the KOG database	185
6.3.6. Light harvesting and carbon fixation	190
6.3.7. Nutrient acquisition and assimilation	195
6.3.7.1. Nitrogen metabolism	195
6.3.7.2. Sulphur metabolism	199
6.3.7.3. Cellular iron utilisation	200
6.4. Discussion	202
6.4.1. Transcriptome statistics and phylogenetics	203
6.4.2. Light harvesting and carbon fixation	205
6.4.3. Nutrient Acquisition and assimilation	209
6.5 Conclusions and further work	213

<b>7. Metatranscriptomics of PPEs along an Atlantic Meridional</b>	
<b>Transect</b>	<b>215</b>
7.1. Introduction	216
7.2. Materials and Methods	220
7.2.1 .Sample collection	220
7.2.2. Flow cytometric sorting of the metatranscriptome seawater sample	221
7.2.3. Extraction of RNA and enrichment of mRNA	221
7.2.4. Sequencing of the metatranscriptomic sample	222
7.2.5. Analysis of 454 reads	222
7.3. Results	222
7.3.1. Pipeline for the extraction and amplification of environmental RNA	222
7.3.2. General analysis of the southern gyre metatranscriptome sequences	224
7.3.3 Phylogenetic analysis of the metatranscriptome	226
7.3.4 Analysis of the AMT18 southern gyre metatranscriptome against the KOG database	227
7.3.5. Light harvesting and carbon fixation	228
7.3.6. Nutrient assimilation and acquisition	230
7.3.7. Utilisation of iron in the metatranscriptome	232
7.4. Discussion	234
7.4.1. Enrichment of mRNA in the metatranscriptome	234
7.4.2. Phylogenetics of the metatranscriptome	235
7.4.3. Functional prediction of contigs	236
7.4.4. Light harvesting and carbon fixation: a metatranscriptomics perspective	236
7.4.5. Metatranscriptomic insights into nutrient acquisition and assimilation	241
7.5. Conclusions and further work	243
<b>8. General discussion and future directions</b>	<b>245</b>
8.1. General discussion	246
8.2. Future directions	251
<b>9. Bibliography</b>	<b>255</b>
<b>10. Appendix</b>	<b>283</b>

## List of Figures

<b>Figure 1.1.</b>	Cartoon depiction of the “biological pump”	4
<b>Figure 1.2.</b>	View of the pelagic food web with the microbial loop being the major path for organic matter flux	6
<b>Figure 1.3.</b>	Abundance and contribution of different groups of phytoplankton to marine carbon fixation	8
<b>Figure 1.4.</b>	Map illustrating the 54 marine provinces described by Longhurst <i>et al.</i> , (2007)	11
<b>Figure 1.5.</b>	Ecological provinces in the NW Atlantic	13
<b>Figure 1.6.</b>	Schematic view of the proposed endosymbiotic events which have given rise to the diversity of plastids	17
<b>Figure 1.7.</b>	Venn diagram showing a comparison of the gene compliments of four sequenced Mamiellophyceae genomes	27
<b>Figure 1.8.</b>	Breakdown of the medium and large clusters of the GOS database into three categories – GOS, known prokaryotic and known eukaryotic	34
<b>Figure 2.1.</b>	RAMMCAAP pipeline for the analysis of metagenomic datasets	65
<b>Figure 3.1.</b>	Cruise track of D326 in the Northeast Atlantic showing the cruise track and station codes	69
<b>Figure 3.2.</b>	Flow cytometry scattergram depicting the PPE size fractions sorted during this study	71
<b>Figure 3.3.</b>	Phylogenetic relationships amongst SOLAS 16S rRNA gene sequences related to the Prymnesiophyceae, obtained from the SOLAS cruise D326	74
<b>Figure 3.4.</b>	Phylogenetic relationships amongst SOLAS 16S rRNA gene sequences related to the Prasinophyceae, obtained from the SOLAS cruise D326	75
<b>Figure 4.1.</b>	Provinces of the Atlantic Ocean as described by Longhurst (1998)	83
<b>Figure 4.2.</b>	AMT18 transect during October/November 2008	85
<b>Figure 4.3.</b>	A) Temperature (°C) and B) salinity (PSU) along the AMT18 cruise track	91
<b>Figure 4.4.</b>	Nutrient profiles along AMT18. A) Phosphate ( $\mu\text{mol l}^{-1}$ ); B) Silicate ( $\mu\text{mol l}^{-1}$ ); C) Nitrite ( $\mu\text{mol l}^{-1}$ ); D) Nitrate ( $\mu\text{mol l}^{-1}$ )	93
<b>Figure 4.5.</b>	Chlorophyll <i>a</i> content along the AMT18. A) Composite of chlorophyll <i>a</i> content from MODIS satellite system during October 2008; B) Measured chlorophyll <i>a</i> content from CTD casts during AMT18	94
<b>Figure 4.6.</b>	Abundance of PPEs (cells $\text{ml}^{-1}$ ) along AMT18	95
<b>Figure 4.7.</b>	Flow cytometry based cell abundance measurements of specific PPE populations along AMT18. A) Abundance of Euk-A (cells $\text{ml}^{-1}$ ); B) Abundance of Euk-B (cells $\text{ml}^{-1}$ )	96
<b>Figure 4.8.</b>	Phylogenetic tree of Prymnesiophyceae sequences with AMT18 sequences showing corresponding OTU numbers	99

<b>Figure 4.9.</b>	The taxonomic composition of PPEs along AMT18 using plastid biased 16S rRNA primers and bulk community DNA	101
<b>Figure 4.10.</b>	Phylogenetic tree of Chlorophyta sequences with AMT18 sequences showing corresponding OTUs	102
<b>Figure 4.11.</b>	Phylogenetic tree of Heterokonta sequences with AMT18 sequences showing corresponding OTUs	104
<b>Figure 4.12.</b>	The taxonomic composition of Euk-A and Euk-B PPE populations along AMT18 at the 55% light level using 16S rRNA gene clone libraries and primers targeted to marine algal plastids	108
<b>Figure 4.13.</b>	The taxonomic composition of Euk-A and Euk-B PPE populations along AMT18 at the DCM using 16S rRNA gene clone libraries and primers targeted to marine algal plastids	108
<b>Figure 4.14.</b>	Dendrogram depicting the composition of OTUs at various locations on AMT18 (CTD number) in the Euk-A and Euk-B PPE populations created using a Euclidean distance matrix	117
<b>Figure 4.15.</b>	Margalef's diversity indices at the 55% light level and DCM along AMT18	119
<b>Figure 4.16.</b>	Composition of AMT 18 stations from OTUs created with nuclear 18S rRNA libraries	122
<b>Figure 4.17.</b>	Phylogenetic tree of nuclear 18S rRNA sequences	124
<b>Figure 4.18.</b>	CCA plot for PPE class distributions along AMT18 in relation to measured environmental variables, using data from clone libraries constructed from bulk community DNA	128
<b>Figure 4.19.</b>	CCA plot for specific PPE OTU distributions along AMT18 in relation to measured environmental variables, using data from clone libraries constructed from bulk community DNA	129
<b>Figure 4.20.</b>	CCA plot for PPE class distributions along AMT18 in relation to measured environmental variables, using data from clone libraries constructed from flow sorted cells	131
<b>Figure 4.21.</b>	CCA plot for specific PPE OTU distributions along AMT18 in relation to measured environmental variables, using data from clone libraries constructed from flow sorted cells	132
<b>Figure 4.22.</b>	The complex interactions which lead to the local diversity in a specific area	147
<b>Figure 5.1.</b>	Phylogenetic tree of 16S rRNA for Chlorophyta species using different culture methods	158
<b>Figure 5.2.</b>	Phylogenetic tree of 16S rRNA for non Chlorophyta species using different culture methods	162
<b>Figure 5.3.</b>	Flow diagram showing possible integrations of techniques to further understand environmentally important species	171
<b>Figure 6.1.</b>	Comparison of the concentration of RNA using two methods of RNA extraction	180



<b>Figure 6.2.</b>	cDNA insert size from the prasinophyte clade VII transcriptome sample	181
<b>Figure 6.3.</b>	Distribution of contig lengths in A) Prasinophyte clade VIIA strain RCC1124 and B) <i>Ochromonas</i> sp. strain CCMP584	182
<b>Figure 6.4.</b>	The binned GC content (%) of A) Prasinophyte clade VIIA strain RCC1124 and B) <i>Ochromonas</i> sp. strain CCMP584	183
<b>Figure 6.5.</b>	Phylogenetic tree indicating the position of the 18S rRNA contig from the clade VII prasinophyte transcriptome	187
<b>Figure 6.6.</b>	Phylogenetic analysis of the top BLASTx hits from the clade VIIA prasinophyte RCC1124 cDNA-derived contigs	188
<b>Figure 6.7.</b>	Phylogenetic analysis of the top BLASTx hits from the marine <i>Ochromonas</i> sp. strain CCMP584 cDNA derived contigs	188
<b>Figure 6.8.</b>	The percentage contribution of KOG classes in the clade VIIA prasinophyte (strain RCC1124) and <i>Ochromonas</i> sp. (strain CCMP584) samples	189
<b>Figure 6.9.</b>	Encoded enzymes and transporters involved in the acquisition and assimilation of nitrogen within either the clade VII prasinophyte	195
<b>Figure 6.10.</b>	Transcripts encoding enzymes of the urea cycle in <i>Ochromonas</i> sp. CCMP584	199
<b>Figure 6.11.</b>	Proposed model for C <sub>4</sub> carbon fixation in <i>Ostreococcus</i> .	209
<b>Figure 7.1.</b>	Environmental genomic analysis of microbial communities from nucleic acids	217
<b>Figure 7.2.</b>	The AMT 18 cruise highlighting the station sampled for metatranscriptomic analysis	220
<b>Figure 7.3.</b>	Flow cytometry scatterplot illustrating the picoeukaryote fraction which was targeted for the metatranscriptome	221
<b>Figure 7.4.</b>	RNA extraction of a variety of cultures	223
<b>Figure 7.5.</b>	PCR product from the amplification of cDNA showing the size range sent for sequencing	224
<b>Figure 7.6.</b>	A) The average length of contigs produced from the southern gyre AMT18 transcriptome sample and B) the binned %GC content of the transcriptome	225
<b>Figure 7.7.</b>	Phylogenetic analysis of the BLASTx top hits from the AMT18 southern gyre metatranscriptome	227
<b>Figure 7.8.</b>	The percentage of contigs assigned to each of the KOG classes	229
<b>Figure 7.9.</b>	Transporters and enzymes associated with nitrogen acquisition and assimilation present in the southern gyre metatranscriptome	230
<b>Figure 7.10.</b>	Proposed pathway for C <sub>4</sub> photosynthesis in marine diatoms	240

## List of Tables

<b>Table 1.1.</b>	Chlorophyll and carotenoid chemotaxonomy of eukaryotic phytoplankton	18
<b>Table 1.2.</b>	Species of algae which have had their genomes published	24
<b>Table 2.1.</b>	Genotype of the three <i>E. coli</i> strains Top10F', JM109 and DH10B T1 Phage Resistant used in this study	40
<b>Table 2.2.</b>	Composition of PPE growth media.	43
<b>Table 2.3.</b>	PCR primer sequences used during this study	49
<b>Table 2.4.</b>	The components used (final concentrations) for the specific PCR reactions	50
<b>Table 3.1.</b>	Number of clones relating to each RFLP type and the resulting contribution to the clone library	73
<b>Table 3.2.</b>	Number of PPE related clones in each library and the corresponding OTUs observed. Number of OTUs predicted using the diversity indicators Chao1 and ACE for each library	76
<b>Table 4.1.</b>	Percentage light levels and the corresponding depths at CTDs along the transect	87
<b>Table 4.2.</b>	The depths at CTDs corresponding to the 55% light level and DCM	87
<b>Table 4.3.</b>	Statistics of clone libraries created from bulk community filters	97
<b>Table 4.4.</b>	Clone library details from flow sorted Euk-A and Euk-B populations along AMT18 including sampling effort and Margalef species richness estimates	107
<b>Table 4.5.</b>	Percentage contribution of classes to clone libraries from flow cytometrically sorted samples	109
<b>Table 4.6.</b>	Numbers of clones relating to specific OTUs in either the Euk-A or Euk-B population plastid 16S rRNA clone libraries constructed from the 55% light level along AMT18	112
<b>Table 4.7.</b>	Numbers of clones relating to OTUs in either the Euk-A or Euk-B population plastid 16S rRNA clone libraries constructed from the DCM along AMT18	114
<b>Table 4.8.</b>	Morisita Horn Similarity Index between the CTDs at the 55% light level for the Euk-A population along AMT18	116

<b>Table 4.9.</b>	Morisita Horn Similarity Index between the CTDs at the DCM for the Euk-A population along AMT18	116
<b>Table 4.10.</b>	Pearson correlation coefficients for environmental variables with Margalef's diversity values at the 55% light level and DCM	120
<b>Table 4.11.</b>	Numbers of clones relating to OTUs in either the Euk-A or Euk-B population 18S rRNA clone libraries constructed from the 55% light level along AMT18	122
<b>Table 4.12.</b>	The number of clones and phylogenetic affiliation of OTUs obtained from the prymnesiophyte specific 18S rRNA gene clone library	126
<b>Table 5.1.</b>	Stations sampled for the collection of starter PPE cultures using direct concentration of natural seawater samples	153
<b>Table 5.2.</b>	Stations sampled for the collection of starter PPE cultures using flow cytometric sorting	154
<b>Table 5.3.</b>	Cultured phytoplankton obtained from AMT 18 using soil extract medium as assessed by sequencing of the plastid 16S rRNA gene	157
<b>Table 5.4.</b>	Taxonomic identification of dominant strains in enrichment cultures from seawater concentrates supplemented with f/2 medium	159
<b>Table 5.5.</b>	Cultured phytoplankton obtained from AMT18 following flow cytometric sorting and enrichment in f/2 medium as assessed by sequencing of the plastid 16S rRNA gene	161
<b>Table 5.6.</b>	Effect of medium type on the composition of PPE enrichment cultures from AMT18	164
<b>Table 6.1.</b>	Phylogenetic analysis of 18S rRNA reads from the <i>Ochromonas</i> transcriptome	185
<b>Table 6.2.</b>	Phylogenetic relationship of 28S rRNA transcripts in the <i>Ochromonas</i> transcriptome	186
<b>Table 6.3.</b>	Transcripts involved in carbon fixation in the clade VII prasinophyte	191
<b>Table 6.4.</b>	Transcripts involved in carbon fixation in <i>Ochromonas</i> transcriptome	192
<b>Table 6.5.</b>	Transcripts relating to carbonic anhydrases in both transcriptomes	193
<b>Table 6.6.</b>	LHC associated transcripts in the clade VII prasinophyte	194

<b>Table 6.7.</b>	Transcripts encoding light harvesting complex proteins in the <i>Ochromonas</i> transcriptome	194
<b>Table 6.8.</b>	Transcripts encoding genes involved in nitrogen assimilation and utilisation in the clade VII prasinophyte transcriptome	196
<b>Table 6.9.</b>	Transcripts encoding genes involved in nitrogen assimilation and utilisation in the clade VII prasinophyte transcriptome	197
<b>Table 6.10.</b>	Transcripts in the <i>Ochromonas</i> transcriptome encoding for components of the urea cycle when analysed against the KEGG database	198
<b>Table 6.11.</b>	Assimilation and transport of sulphur in the clade VII prasinophyte transcriptome	200
<b>Table 6.12.</b>	Assimilation and transport of sulphur in the <i>Ochromonas</i> transcriptome	200
<b>Table 6.13.</b>	Iron utilisation in the clade VII prasinophyte transcriptome	201
<b>Table 6.14.</b>	Iron utilisation in the <i>Ochromonas</i> transcriptome	201
<b>Table 6.15.</b>	Transcripts involved in the acquisition of iron in the two transcriptomes	202
<b>Table 7.1.</b>	Phylogenetic affiliation of eukaryotic rRNA transcripts in the metatranscriptome	225
<b>Table 7.2.</b>	Transcripts involved in carbon fixation and the phylogenetic affiliation	231
<b>Table 7.3.</b>	Blast results for transcripts relating to photoreceptors and the phylogenetic affiliation	232
<b>Table 7.4.</b>	Transcripts identified in the metabolism and transport of nitrogen	233
<b>Table 7.5.</b>	Utilisation and transport of iron in the AMT 18 metatranscriptome	234



## Acknowledgements

Firstly I would like to thank my supervisors Prof. David Scanlan and Prof. Mikhail Zubkov for the help which they have given me during the four years of my PhD as well as my committee members Prof. Colin Murrell, Dr. Kevin Purdy and Prof. Bruno Frenguelli.

Secondly I would like to thank the past and current members of MicroII with whom I have shared a lab. I would especially like to thank Sophie Mazard, Martin Ostrowski and Ludwig Jardillier for their help in solving the various issues which developed. My thanks are also extended to my bay members Viktoria Eriksson, Andrew Millard and Jackie Chan for keeping me close to sanity during the long hours in the lab. Thanks must also go to Carolina Grob, Cecile Lepere and Sophie Mazard for their proof reading expertise. I would also like to thank Ross Holland at the National Oceanographic Centre, Southampton, for his help with the flow cytometry work presented here.

I would also like to express my gratitude to the crew of the RRS James Clark Ross, who successfully managed to get us from the UK to the Falkland Islands during AMT 18. In particular I would like to thank Nerys Lewis, the doctor on the cruise for patching me up and making sure I arrived back in one piece. Also thanks to the penguins of the Falkland Islands being a welcome sight after 7 weeks at sea.

Finally I would like to express my thanks to my parents and sister for their support during my PhD.

## **Declaration**

I declare the work presented in this thesis was conducted by me under the direct supervision of Prof. David J. Scanlan, with exception of those instances where the contribution of others has been specifically acknowledged. None of the work presented has been previously submitted for any other degree

John Pearman

## Summary

Photosynthetic picoeukaryotes (PPEs), defined here as single celled organisms <3  $\mu\text{m}$  in diameter, are significant contributors to primary production. Until recently, marine PPEs had received relatively little research attention in contrast to the more numerous picocyanobacteria. Molecular studies have now started to reveal the diversity of this group, using both the nuclear 18S rRNA gene and the plastid-targeted 16S rRNA gene as taxonomic markers. The latter marker has the advantage of directly targeting the PPE community, counteracting the problem of heterotrophic sequences dominating clone libraries. As well as PCR based molecular approaches, genomic studies of PPEs are starting to reveal the metabolic capabilities of these organisms.

In this thesis, taxonomic information obtained on two flow-sorted PPE populations (Euk-A and Euk-B) showed that pico-prymnesiophytes, largely representing lineages with no close cultured counterpart, dominated the Euk-A and Euk-B libraries (54 and 58%, respectively) in tropical and sub-tropical waters of the Atlantic Ocean. Radiotracer work performed elsewhere had shown these PPE groups contribute up to 19% and 38% (Euk-A and Euk-B, respectively) to total  $\text{CO}_2$  fixation, demonstrating the importance of these PPE groups in marine carbon cycling.

To further assess the taxonomic composition and distribution of these Euk-A and Euk-B PPE populations at the ocean-basin scale, clone libraries were constructed along an Atlantic Meridional Transect (AMT18). Major components of these flow cytometry sorted PPE populations were Prymnesiophyceae and Chrysophyceae using plastid markers, or Prasinophyceae and Dinophyceae (nuclear markers) including several lineages with no cultured counterparts. In surface waters a latitudinal diversity gradient was observed with a peak in PPE diversity found in the equatorial region. Distribution patterns of specific PPE groups and OTUs were subsequently correlated with measured environmental parameters, although most of the variation in PPE diversity was not explained by the measured variables.

Attempts were undertaken to obtain into culture novel PPEs, especially those representative of oligotrophic regions. However, the majority of isolates obtained were related to *Prasinoderma* or *Chlorella* which are cosmopolitan, fast-growing genera. Even so, some isolates more relevant of open ocean environments were obtained, including a clade VIIA prasinophyte and a *Pelagomonas* sp.

Transcriptomics was used to further assess the functional potential of specific PPE populations, firstly in cultures using both an *Ochromonas* sp. and a prasinophyte as being representative of organisms present along AMT18. This approach revealed a  $\text{C}_4$  carbon concentrating mechanism in the clade VIIA prasinophyte and enzymes required for a functioning urea cycle in the *Ochromonas* sp. A pipeline was also developed to undertake a metatranscriptomic approach on a flow cytometrically sorted PPE population from the south Atlantic gyre. This approach revealed a diatom-like  $\text{C}_4$  carbon concentrating system in the metatranscriptome.

Overall, this thesis has given new insights into the diversity of specific PPE groups at the ocean basin-scale, developed a new pipeline for the transcriptomic analysis of PPEs both in culture and in the environment, and in so doing has provided new information on the functional potential of these important photosynthetic organisms.

## Abbreviations

ABC	ATP-binding cassette
ACC	Antarctic Circumpolar Current
ACE	abundance-based coverage estimator
ALSK	Alaska Downwelling Coastal Province
AMD	Acid Mine Drainage
Amp	ampicillin
AMT	Atlantic Meridional Transect
AMT	Ammonia Transporter Channel
ANTA	Antarctic Province
APLR	Austral Polar Province
ARAB	Northwestern Arabian Upwelling Province
ARCH	Archipelagic Deep Basins Province
ARCT	Atlantic Arctic Province
aRNA	amplified ribonucleic acid
ATP	adenosine triphosphate
AUSE	East Australian Coastal Province
AUSW	Australia-Indonesia Coastal Province
BENG	Benguela Current Coastal Province
BERS	North Pacific Epicontinental Se Province
BLAST	Basic Local Alignment Search Tool
bp	base pair
BPLR	Boreal Polar Province
BRAZ	Brazil Current Coastal Province
BUT	butanoyloxyfucoxanthin
CA	carbonic anhydrase
CAMERA	Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis
cAMP	cyclic adenosine monophosphate
CAMR	Central American Coastal Province
CARB	Caribbean Province
CCA	Canonical correspondence analysis
CCAL	California Upwelling Coastal Province
CCM	carbon concentrating mechanism
CCMP	Provasoli-Guillard National Centre for Culture of Marine Phytoplankton
CD-HIT	Cluster Database-High Identity with Tolerance
cDNA	complementary DNA
CHIL	Chile-Peru Current Coastal Province
CHIN	China Sea Coastal Province
CHLa	chlorophyll <i>a</i>
CNRY	Eastern Canary Coastal Province
CTD	conductivity, temperature, depth
DASS	dual sodium/sulphate symporter
DCM	Deep chlorophyll maximum
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DOM	dissolved organic material
DTT	dithiothreitol



EAFR	Eastern Africa Coastal Province
EDTA	Ethylenediaminetetraacetic acid
EST	expressed sequence tag
EtBr	ethidium Bromide
ETRA	Eastern Tropical Atlantic Province
FET3p	copper (Cu)-containing ferroxidase
FISH	fluorescence in situ hybridisation
FLKD	Southwest Atlantic Shelves Province
FUC	fucoxanthin
GFST	Gulf Stream Province
GMD	gel microdroplets
GOS	Global Ocean Sampling
GUIA	Guianas Current Coastal Province
GUIN	Guinea Current Coastal Province
HEX	19' -hexanoyloxyfucoxanthin
HMM	hidden Markov model
HNLC	high nutrient low chlorophyll
HPLC	high performance liquid chromatography
INDE	Eastern India Coastal Province
INDW	Western India Coastal Province
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
ISSG	Indian South Subtropical Gyre Province
kb	kilobase
KEGG	Kyoto Encyclopedia of Genes and Genomes
KOG	cluster of orthologous groups for eukaryotic complete genomes
KURO	Kuroshio Current Province
LB	Luria Broth
LHC	light harvesting complex
LiDS	Lithium dodecyl sulfate
LSU	large subunit
MAST	marine stramenopile
MDH	malate dehydrogenase
ME	malic enzyme
MEDI	Mediterranean Sea, Black Sea Province
MEGAN	Metagenome Analyze
MFS	Major Facilitator Superfamily
MONS	Indian Monsoon Gyres Province
mRNA	Messenger ribonucleic acid
NAD	nicotinamide adenine dinucleotide
NASE	North Atlantic Subtropical Gyral East Province
NASW	North Atlantic Subtropical Gyral West Province
NATR	North Atlantic Tropical Gyre Province
NCBI	National Center for Biotechnology Information
NECS	Northeast Atlantic Shelves Province
NERC	National Environmental Research Council
NEWZ	New Zealand Coastal Province
NPPF	North Pacific Transition Zone Province
NPSW	North Pacific Subtropical Gyre West Province
NPTG	North Pacific Equatorial Countercurrent Province
nr	non redundant

NW	Northwest
NWCS	Northwest Atlantic Shelves Province
OAA	oxaloacetate
ODV	Ocean Data View
ORF	open reading frame
OTU	operational taxonomical unit
PAP	Porcupine Abyssal Plain
PAR	Photosynthetically active radiation
PAS	Per – period circadian protein Arnt – aryl hydrocarbon receptor nuclear translocator protein Sim – single-minded protein
PCR	Polymerase chain reaction
PEP	phosphoenolpyruvate
PEPCase	phosphoenolpyruvate carboxylase
PEPCKase	phosphoenolpyruvate carboxykinase
PFA	paraformaldehyde
PFAM	protein family database
PLN	proximate limiting nutrient
PPDK	pyruvate orthophosphate dikinase
PPE	Photosynthetic picoeukaryotes
PSAE	Pacific Subarctic Gyre East Province
PSAW	Pacific Subarctic Gyre West Province
<i>psbA</i>	gene encoding the D1 protein of photosystem II
<i>psbS</i>	photosystem II subunit
PSU	practical salinity unit
PYC	pyruvate carboxylase
RAMMCAP	Rapid Analysis of Multiple Metagenomes with a Clustering and Annotation Pipeline
RCC	Roscoff Culture Collection
REDS	Red Sea, Persian Gulf Province
RFLP	restriction fragment length polymorphism
rRNA	ribosomal ribonucleic acid
RRS	Royal Research Ship
RuBisCO	ribulose-1,5-bisphosphate carboxylase oxygenase
SANT	Subantarctic Water Ring Province
SARC	Atlantic Subarctic Province
SATL	South Atlantic Gyral Province
SDS	sodium dodecyl sulphate
SOD	superoxide dismutase
SOLAS	Surface Ocean - Lower Atmosphere Study
SPSG	South Pacific Subtropical Gyre Province
SSC	side scatter
SSD	Sargasso Sea Database
SSTC	South Subtropical Convergence Province
SSU	small subunit
SulP	sulphate transporter
SUND	Sunda-Arafura Shelves Province
TASM	Tasman Sea Province
TIGRfam	curated multiple sequence alignments for protein sequence classification
ULN	ultimate limiting nutrient

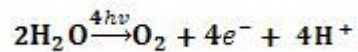
UV	Ultraviolet
VPA	Variation Partitioning Analysis
WARM	Western Pacific Warm Pool Province
WTRA	Western Tropical Atlantic Province
XGal	5-bromo-4-chloro-3-indolyl- beta-D-galactopyranoside

# **Chapter 1**

## **Introduction**

### 1.1 Evolution of photosynthesis

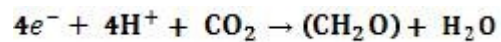
Until ~2500 million years ago much of biology was dependent on anoxygenic photosynthesis, with H<sub>2</sub>S, NH<sub>3</sub>, organic acids and Fe<sup>2+</sup> utilised as hydrogen/electron donors (Barber, 2008). Anoxygenic photosynthesis is undertaken by four lineages of bacteria: purple bacteria, green-sulphur bacteria, green non-sulphur bacteria and Heliobacteria (Xiong *et al.*, 2000). Around 2500 million years ago, an ‘oxygenic’ photosynthetic apparatus evolved which could utilise solar power to split water in order to provide protons/electrons:



It is believed that oxygenic photosynthesis evolved from bacterial anoxygenic photosynthesis. An analysis of photosynthetic genes by Xiong *et al.*, (2000) suggested that Heliobacteria were the closest relatives to those organisms undertaking oxygenic photosynthesis. In extant cyanobacteria over 100 genes are involved in oxygenic photosynthesis (Falkowski and Godfrey, 2008). The lateral transfer of subsections of the photosynthetic apparatus appears to be insufficient in establishing the pathway in other prokaryotic clades (Falkowski and Godfrey, 2008). Consequently, unlike other core prokaryotic metabolic pathways, oxygenic photosynthesis is limited to the cyanobacterial clade. However, the wholesale engulfment of a cyanobacterium by a heterotrophic eukaryote, led to the formation of plastids (see section 1.6), which enabled eukaryotes to be capable of oxygenic photosynthesis.

The emergence of oxygenic photosynthesis had a profound effect on biology and the Earth’s atmosphere. The by-product of the splitting of H<sub>2</sub>O was molecular oxygen.

With the release of oxygen the world was slowly converted from an anaerobic environment into an aerobic one. The resulting electrons and protons, from the splitting of water, enabled the transformation of inorganic carbon into sugars and other organic molecules (e.g. see Barber, 2008):



The availability of oxygen led to an increased efficiency in metabolism, as aerobic respiration provides in the region of 20 times more cellular energy than anaerobic respiration (Barber, 2008). Furthermore, the establishment of an ozone layer provided a shield against the dangers of UV radiation and enabled the colonisation of terrestrial environments. When organisms evolved the ability to split water, it led to the evolution of eukaryotic cells and multicellular life was able to prosper and diversify, as exemplified by the fossil record and the number of extant species currently present on the planet (Barber, 2008).

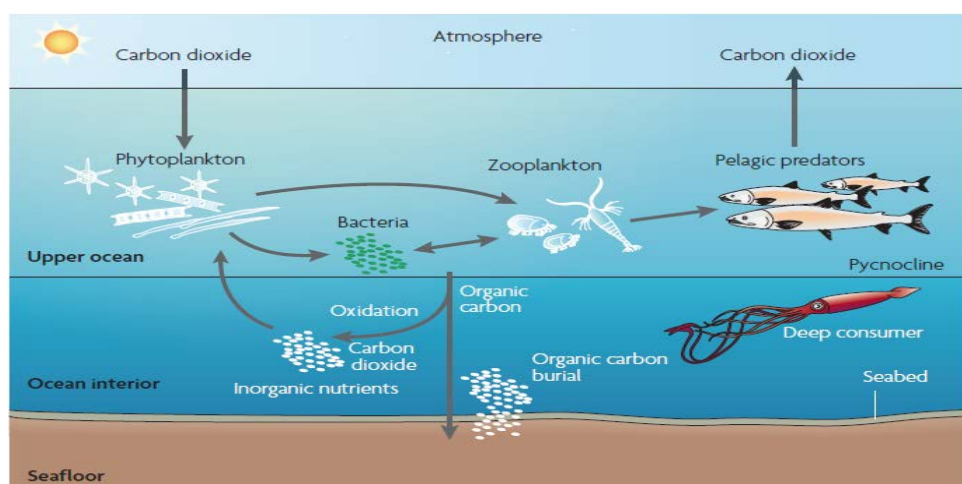
## 1.2 Global carbon fixation

Carbon is cycled around three major reservoirs in the global system and undergoes a variety of changes in chemical forms as it does so. The three reservoirs are the ocean, atmosphere and terrestrial system (Falkowski *et al.*, 2000). The largest reservoir for carbon is in the ocean, whilst the smallest reservoir is the atmosphere. Since the industrial revolution, the burning of fossil fuels has released significantly more carbon into the atmosphere from the geological store.

The marine environment has a major role in determining the concentration of CO<sub>2</sub> in the atmosphere. The concentration is determined via physical processes (mixing and circulation), chemical processes (carbon chemistry and buffering effects) and

biological processes (production and decomposition of organic matter) (Post *et al.*, 1990). Biological processes are particularly important in the marine environment. Carbon is stored in three forms: dissolved inorganic carbon (consisting of dissolved  $\text{CO}_2$ , the bicarbonate and carbonate ions  $\text{HCO}_3^-$  and  $\text{CO}_3^{2-}$ ), dissolved organic carbon and particulate organic carbon (consisting of live organisms plus fragments of dead plants and animals) (Post *et al.*, 1990). Winds across the ocean surface mean that  $\text{CO}_2$  is rapidly exchanged across the sea-air interface. This results in an approximate equilibrium between the partial pressures of  $\text{CO}_2$  in the atmosphere and surface waters (Post *et al.*, 1990).

In order for  $\text{CO}_2$  to be taken up into surface waters carbon needs to be fixed into organic matter and eventually transported into the deep ocean: the combined processes of fixation of inorganic carbon into organic matter during photosynthesis, its transformation by food web processes, transport, together with physical mixing and gravitational setting are known as the “biological pump” (Figure 1.1) (Ducklow *et al.*, 2001).



**Figure 1.1. Cartoon depiction of the “biological pump” (taken from Falkowski and Oliver, 2007).**

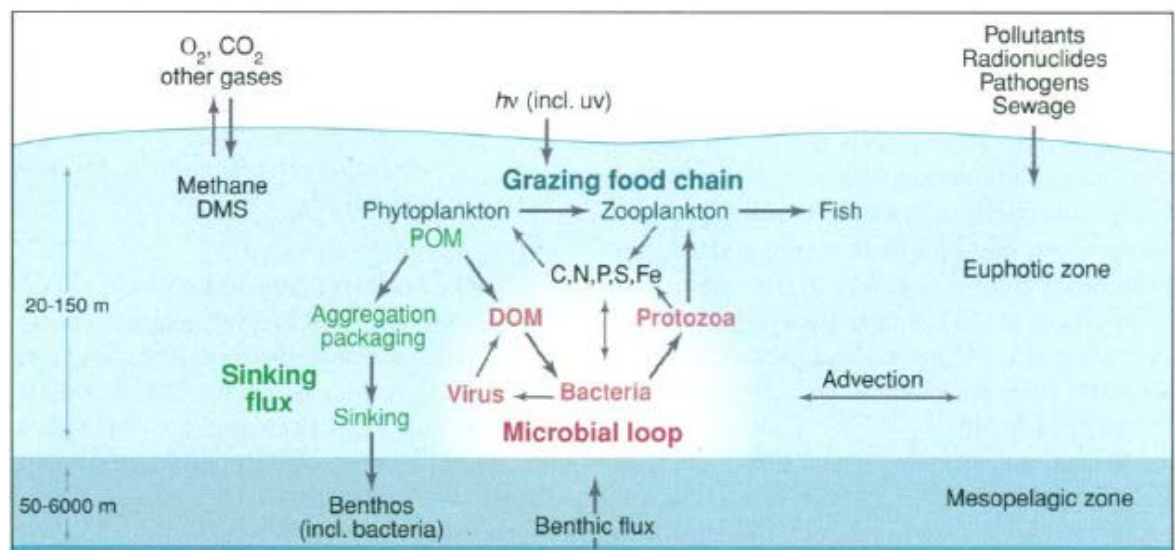
Biochemical production of organic material is almost completely accounted for by oxygenic photosynthesis (Field *et al.*, 1998). On a global scale, about 50% of carbon fixation is attributed to oceanic environments, with the remainder being undertaken in terrestrial habitats (Field *et al.*, 1998). In terrestrial environments, vascular plants dominate primary production, with only a minor contribution from microorganisms. In the aquatic system free living photosynthetic microbes are the major primary producers (Field *et al.*, 1998). The majority of carbon fixation in the oceanic environment is undertaken by microbial phytoplankton, with only 1-5% being attributed to benthic organisms such as macroalgae, cyanobacteria and seagrasses (Gadd and Raven, 2010). Size fractionated analysis of chlorophyll *a* concentrations have found that cells <2  $\mu\text{m}$  had the greatest contribution to phytoplankton biomass in open ocean environments (Teira *et al.*, 2005; Perez *et al.*, 2006).

Much of the photosynthetically produced organic matter is now believed to enter into the 'microbial loop' via the uptake of dissolved organic material (DOM) by bacteria (Figure 1.2) (Azam, 1998). Viruses have been shown to have an impact on the microbial loop. The lysis of cells promotes the recycling of nutrients by bacteria in the water column and limits the transfer of organic matter to higher trophic levels (Middelboe and Lyck, 2002).

A small proportion of the photosynthetically derived organic matter enters the marine food web, with the microbial phytoplankton being consumed by zooplankton and then larger organisms. These larger organisms play a vital part in the transport of carbon into the deep ocean. The passive sinking of faecal pellets and the active diel migration of organisms affect the flux of organic material into the deep ocean (Ducklow *et al.*, 2001). Many species of zooplankton live below the euphotic zone



during the day and undergo a vertical migration into surface waters to feed at night. This transport of surface ingested organic material can increase the magnitude of carbon export into the deep ocean (Ducklow *et al.*, 2001). A minor fraction of the carbon exported into the deep ocean will be deposited on the sea floor, where it can be incorporated into the geological store (Post *et al.*, 1990).



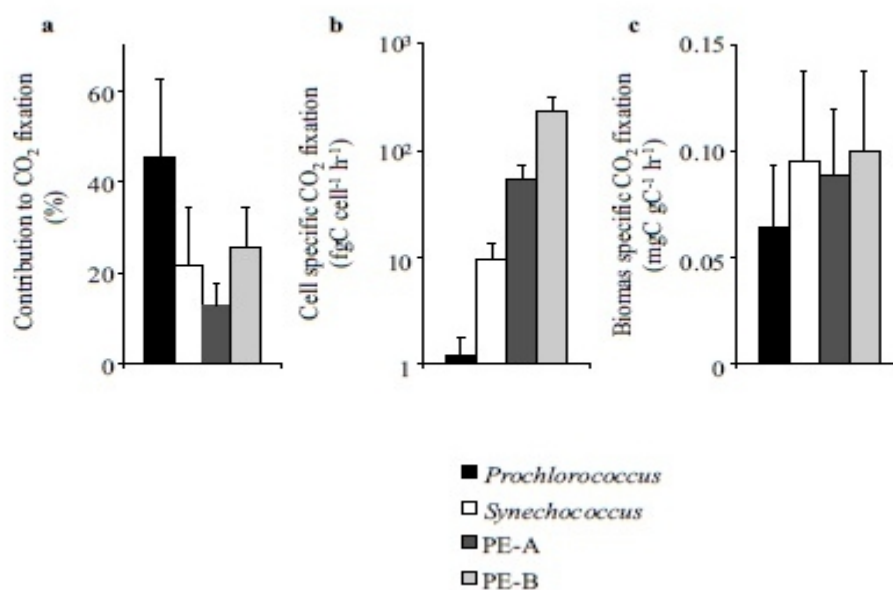
**Figure 1.2. View of the pelagic food web with the microbial loop being the major path for organic matter flux (Azam, 1998)**

### 1.3. Primary production in the marine system

Primary production is dominated in the oceanic environment by planktonic organisms, with only a minor contribution being made by the macroalgae in coastal regions (Field *et al.*, 1998). Size fractionation studies have shown that phytoplankton in the 2-5  $\mu\text{m}$  size fraction play an important role in marine primary production (Perez *et al.*, 2006; Teira *et al.*, 2005). The 2-5  $\mu\text{m}$  size fraction contains both prokaryotic and eukaryotic components. It is well known that the prokaryotic component of the picophytoplankton comprises a wide range of diversity within only two genera: *Synechococcus* and *Prochlorococcus* (see e.g. Scanlan and West, 2002).

However, photosynthetic picoeukaryotes (PPEs), defined here as organisms  $\leq 3 \mu\text{m}$  in size to reflect the fact that recent field studies have often used  $3 \mu\text{m}$  filters to separate small plankton from larger cells (see Vaultot *et al.*, 2008), are much more diverse than their prokaryotic counterparts, with several algal classes identified, including members of the classes Prymnesiophyceae, Prasinophyceae and Pelagophyceae (Slapeta *et al.*, 2006; Vaultot *et al.*, 2008). The prokaryotic component tends to be numerically dominant, particularly in oligotrophic open ocean environments e.g. in one North Atlantic study average concentrations of *Prochlorococcus* were  $1.6 \times 10^8 \text{ cells l}^{-1}$ , *Synechococcus*  $2.35 \times 10^6 \text{ cells l}^{-1}$  whilst PPEs comprised only  $1.0 \times 10^6 \text{ cells l}^{-1}$  (Li, 1994). Thus, *Prochlorococcus* abundances are generally higher in more stratified waters, while *Synechococcus* and PPEs show higher abundances when mixing prevails (Shalapyonok *et al.*, 2001; Grob *et al.*, 2007). However, it was not until the combination of direct flow cytometric sorting experiments with radiotracer experiments that a greater understanding of carbon fixation by these small phytoplankton could be achieved. Despite the numerical domination of the pico-sized phytoplankton by the prokaryotic members, it has been shown that the eukaryotic component makes a significant contribution to primary production. Radiotracer experiments on four flow cytometry sorted groups (*Prochlorococcus*, *Synechococcus*, Euk-A PPEs (comprising cells on average ca  $1.8 \mu\text{m}$  in size) and Euk-B PPEs (comprising cells on average  $2.8 \mu\text{m}$  in size)), found that the rate of cell-specific  $\text{CO}_2$  fixation in the prokaryotic genera was lower than that in the eukaryotic size fractions (Jardillier *et al.*, 2010). Despite the Euk-B group having the lowest abundance, its higher rate of  $\text{CO}_2$  fixation meant that it contributed to, on average, 25% of primary production (Figure 1.3) (Jardillier *et al.*, 2010). This is a similar finding to that of Li (1994), who found that the

contribution of eukaryotic size fractions was disproportionate to their abundance in the environment. In the coastal environment, Worden *et al.*, (2004) found that picoeukaryotes could account up to 75% of the carbon fixation. Despite their ecological importance, the diversity and distribution of PPEs in the marine environment has received far less attention than their prokaryote counterparts although this is starting to change.



**Figure 1.3. Abundance and contribution of different groups of phytoplankton to marine CO<sub>2</sub> fixation (adapted from Jardillier *et al.*, 2010). PE-A and PE-B refer to the Euk-A and Euk-B PPE groups mentioned in the text.**

#### 1.4. Control of primary production

In the marine environment nutrient limitation has a potentially important effect on primary production. Nitrogen and phosphorus are the two main macronutrients which have been proposed to limit primary production (Tyrrell, 1999). Nitrate and phosphate both increase with depth in a ratio of ~15:1 as first realised by Redfield

(Tyrell, 1999). There are two definitions of limiting nutrients: i) The proximate limiting nutrient (PLN) represents the local limiting nutrient with enhancement of primary production occurring within hours of enrichment ii) The ultimate limiting nutrient (ULN) whose supply forces total system productivity over long timescales (Tyrell, 1999). In the North Atlantic it has been suggested that nitrogen is the proximate limiting nutrient although for some phytoplankton groups co-limitation by both nitrogen and phosphorus is apparent (Davey *et al.*, 2008; Moore *et al.*, 2008).

Since the early 1990s the role of iron as a limiting nutrient has received much attention (Martin 1990), particularly its impact on marine nitrogen fixation and thus the amount of new nitrogen available for use by other phytoplankton (Falkowski *et al.*, 1998). Iron inputs into marine systems are mainly via aeolian dust causing iron limitation of phytoplankton photosynthesis in much of the Pacific (Behrenfeld and Kolber, 1999) but not in the Atlantic Ocean (Jickells *et al.*, 2005). In these so-called high-nutrient low-chlorophyll regions (HNLC), such as the subarctic Pacific and Southern Oceans, iron availability appears to be the primary factor limiting production (Hall and Safi, 2001). The low levels of bioavailable iron prevent the complete utilisation of ambient nitrate levels (Hutchins and Bruland, 1998). Thus, about 50% of the nitrogen and phosphorus in the euphotic zone of the Southern Ocean is returned to the deep ocean as the Antarctic Circumpolar Current (ACC) subducts underneath nutrient poor waters to its north (Falkowski *et al.*, 1998). Iron fertilisation experiments in the Southern Ocean have shown strong initial responses from the phytoplankton, with a doubling in cell abundance in the first seven days after fertilisation before returning to pre fertilisation levels (Hall and Safi, 2001; Boyd *et al.*, 2007). In these experiments it has been suggested that grazing controls

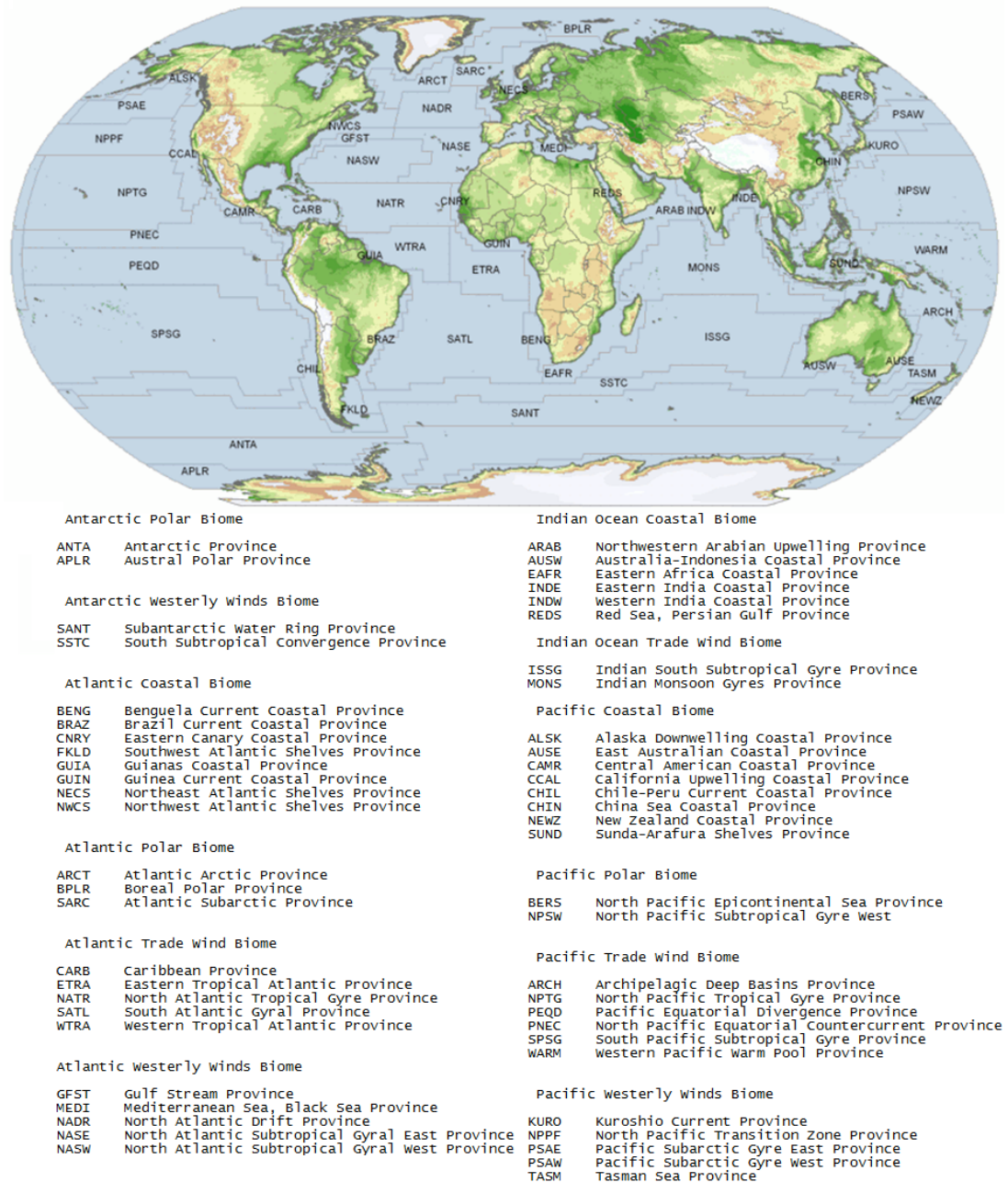
the increase in phytoplankton cell numbers, which in turn damps down levels of primary production. In other regions, such as the equatorial Pacific, where there is an excess of nitrate in surface waters, it is believed that the standing crop of phytoplankton (and hence primary production) is kept low by grazing pressure (Cullen *et al.*, 1992). However, the ultimate limitation in the system may still be iron. Small phytoplankton may dominate the region because they are better scavengers for iron and thus exclude larger diatom species. However, small phytoplankton are more susceptible to grazing pressure compared to the larger diatoms (Cullen *et al.*, 1992). Indeed, iron fertilisation experiments have shown that increased levels of iron leads to a switch from a system dominated by picophytoplankton to one consisting of blooms of diatoms (Hutchins and Bruland, 1998).

### **1.5. Biogeography of the marine environment**

Various studies of marine species, from bacterioplankton to whales, have investigated the distribution of these organisms. These investigations have been aided by the delineation of the marine system into provinces with similar physical and biological properties (Longhurst, 2007).

In the terrestrial system the concept of biomes has long been held but in the oceanic system, the idea of partitioning into biogeochemical regions is a more recent development (Platt and Sathyendranath, 1988). Initially, the oceanic system was divided up into four biomes (Polar, Westerlies, Trade winds and Coastal) which are defined by the dominant oceanographic processes that determine vertical density structure of the water column (Longhurst, 2007). However, the characterisation of the four biomes failed to predict regional structures in the pelagic ecosystem. This

led to the partitioning of the biomes into a total of 56 provinces (Figure 1.4) (Longhurst, 1995).



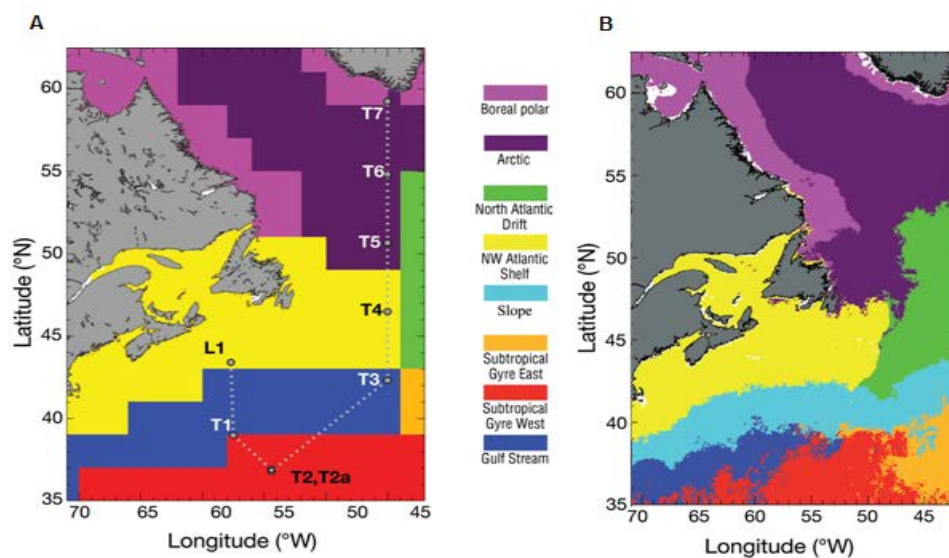
**Figure 1.4.** Map illustrating the 56 marine provinces described by Longhurst, (2007). Available online at <http://www.vliz.be/vmdcdata/vlimar/> Consulted on 2011-07-29.

A limitation in the provinces determined by Longhurst (1995) is the static definition of the boundaries. The determinants of terrestrial biomes are climate and geology, whereas the pelagic ecosystem is almost totally independent of geology (Platt and Sathyendranath, 1999). With the primary determinant being climate, ocean circulation and water mass location varies with seasonal cycles, in response to atmospheric events (Devred *et al.*, 2007). The use of satellite data for sea surface temperature (SST) and chlorophyll *a* concentration has enabled the development of a dynamic method for the ecological partitioning of the ocean in real time. Devred *et al.*, (2007), determined water masses with *in situ* measurements to establish the similarity of dynamic boundaries to the static boundaries determined by Longhurst (1995). Measurements in the NW Atlantic showed that, broadly, the dynamic boundaries were consistent with the static boundaries previously determined (Figure 1.5). However, there were differences in the boundaries with changing seasons; the availability of real time data to assess the ecological provinces, will allow for an improved extrapolation of results, at a given station, to the large scale (Devred *et al.*, 2007).

### **1.6. Plastid evolution as a basis for photosynthetic eukaryote diversity**

The complex acquisition of plastids may have led to the diversification of photosynthetic eukaryotes. The origin of photosynthetic plastids in eukaryotes is believed to have come about through the engulfment of a cyanobacterium by an unknown heterotrophic protist about 1.8 billion years ago (Tirichine and Bowler, 2011). This primary endosymbiotic event led to three chloroplast lineages: Glaucophyta (which have retained the peptidoglycan wall of cyanobacterial origin), Rhodophyta (red algae) and Chlorophyta (green algae and land plants) (Bhattacharya

*et al.*, 2004). Studies based on the highly conserved nature of the photosynthetic apparatus and the sequence similarity between over 100 genes remaining in the chloroplast, suggest that these three lineages are monophyletic (Rodriguez-Ezpeleta *et al.*, 2005). However, due to the time lapse since divergence, the order in which the three lineages branched off has not been resolved (Green, 2011). The plastids found in the primary endosymbiotic lineages have reduced genomes which do not encode all the necessary genes required for plastid function and maintenance. Genome reduction has occurred with the loss of genes not required in the new cellular environment as well as the horizontal transfer of genes to the host nucleus and recruitment of pre existing nuclear genes (Moreira and Philippe, 2001).



**Figure 1.5. Ecological provinces in the NW Atlantic as described by A) Longhurst (1995) and B) using ocean colour radiometry (adapted from Devred *et al.*, 2007).**

The diversity of photosynthetic eukaryotes is complicated by a complex series of secondary and tertiary endosymbiotic events (Figure 1.6). The possibility of



secondary endosymbiosis was first noted, in the alga *Euglena*, back in the 1970s (Gibbs, 1978). Secondary endosymbiosis involves the incorporation of a photosynthetic eukaryote (derived from a primary endosymbiotic event) by a heterotrophic eukaryote (Archibald, 2009a). This type of eukaryote–eukaryote endosymbiosis explains the vast majority of algal diversity (Bhattacharya *et al.*, 2004). Secondary endosymbiotic events can be identified by the presence of three or four bounding membranes around the plastid (Bhattacharya *et al.*, 2004). In general it is believed that secondary endosymbiosis occurs via endocytosis, so the alga is taken up into a vacuole derived from the host endomembrane. With the reduction of the engulfed cell, four membranes would surround the plastid: the host endomembrane, the plasma membrane of the engulfed cell and the two membranes of the primary plastid (Keeling, 2010). Two secondary endosymbiotic events occurred with the engulfment of a member of the green lineage. These secondary endosymbiotic events resulted in Chlorarachniophyta, unicellular marine amoebae containing a green plastid, and Euglenophyta, algae typical of organic-rich freshwater habitats (Bhattacharya *et al.*, 2004).

Red algal derived plastids are found in at least six algal lineages, including cryptophytes, haptophytes, plastid-bearing stramenopiles (e.g. diatoms), apicomplexans, dinoflagellates and *Chromera velia* (Archibald, 2009a). The secondary spread of plastids, especially from the red algal lineage, had a profound impact on eukaryotic diversity, evolution and global ecology. Several lineages have grown to dominate primary production in their environments and it has been estimated that red algal derived secondary endosymbiosis alone accounts for 50% of the described protist species (Keeling, 2010). The case for the secondary spread of

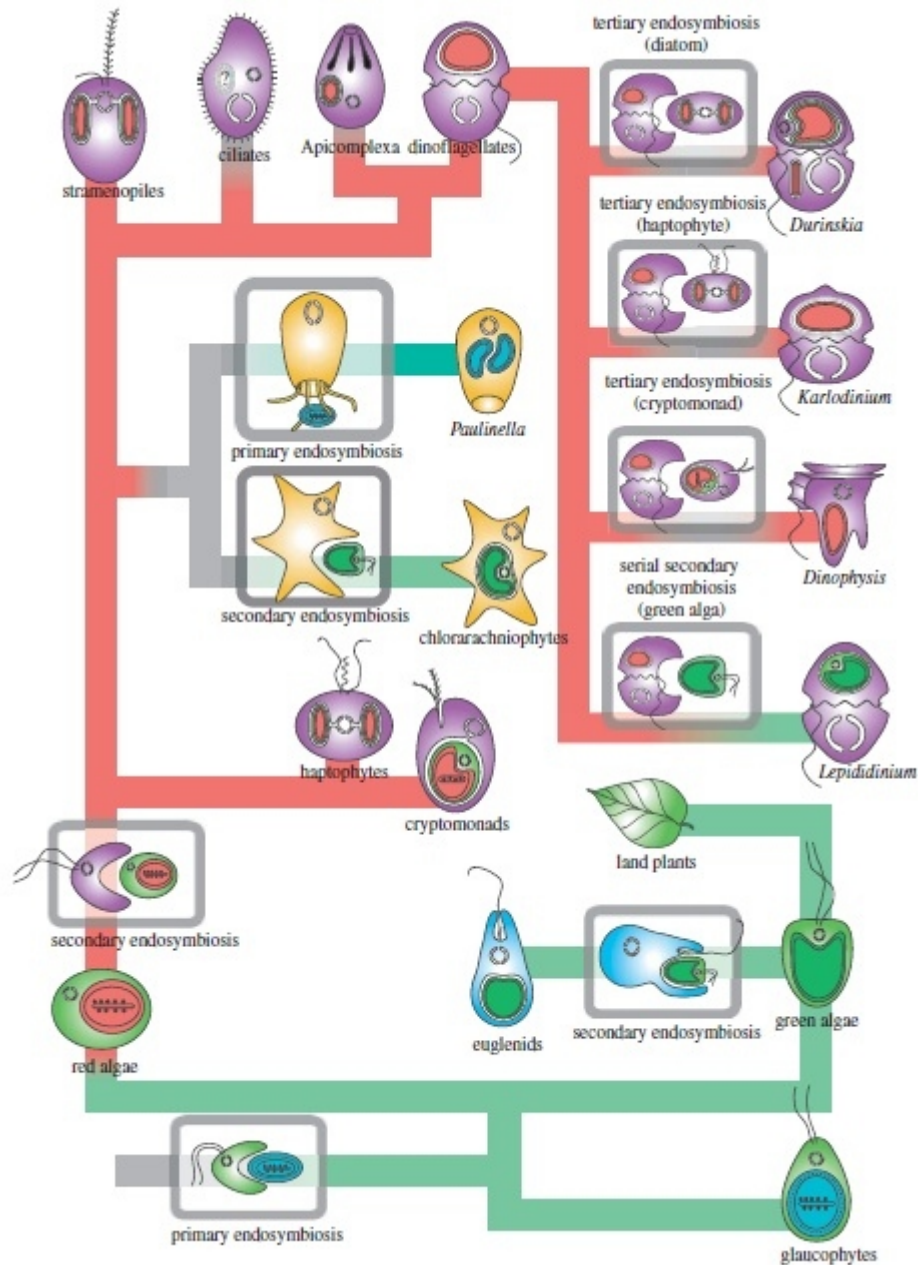
red algal plastids in cryptophytes, haptophytes, stramenopiles, apicomplexans is clear cut (Archibald, 2009a). However, in the dinoflagellates, tertiary endosymbiosis has added a further level of complexity to plastid evolution. The most common type of plastid contained in dinoflagellates contains the carotenoid peridinin (Reyes-Prieto *et al.*, 2007). Several lineages of dinoflagellates no longer have a functioning plastid from the original secondary endosymbiosis. Instead primary production is carried out by a new plastid derived from another lineage. Studies of dinoflagellates have shown that they have acquired tertiary plastids from cryptomonads (Schnepf and Elbrächter, 1988), haptophytes (Tengs *et al.*, 2000) and diatoms (Chesnick *et al.*, 1997), which are all from the red algal lineage. Dinoflagellates have been shown to have at least two different haptophyte tertiary endosymbionts (one in *Karlodinium* and a different one in *Karenia*) (Gast *et al.*, 2007) and at least three different diatom endosymbionts. Furthermore, dinoflagellates have been shown to have acquired a new plastid from the green algae lineage, a case of serial secondary endosymbiosis (Watanabe *et al.*, 1990).

### **1.7. Diversity of photosynthetic picoeukaryotes (PPEs)**

For many years the only way to investigate the diversity of phytoplankton cells was through the use of light microscopy. In the open ocean repeated convergent evolution, towards a small coccoid morphology, has meant that assessing diversity by microscopy has its limitations (Potter *et al.*, 1997). These limitations, as well as a lack of available organisms in culture, has meant that only about 70 species of PPEs have been described (having a size  $\leq 3 \mu\text{m}$ ) (Le Gall *et al.*, 2008). Of these, around 30 species can be considered picoplanktonic based on formal classification (i.e. with a size  $\leq 2 \mu\text{m}$ ) (Vaulot *et al.*, 2008).

In the late 1970s epifluorescence microscopy was used to investigate microbes in the marine environment and led to the conceptualization of the microbial loop (Azam *et al.*, 1983). Epifluorescence microscopy relies on the emission of light by cellular compounds (e.g. pigments such as chlorophyll *a*, or by stains for specific components such as DNA) and has been used to investigate the distribution and diversity of PPEs in the marine environment (Murphy and Haugen, 1985). This technique can be used to identify different cells discriminated on the basis of pigment content, as well as the use of stains such as primulin, which stains phospholipids, to identify features such as flagella (Vaulot *et al.*, 2008).

The biomass of marine phytoplankton has long been measured by investigating the concentration of chlorophyll *a*. The development of high performance liquid chromatography (HPLC) methodologies has facilitated more accurate measurements of chlorophyll *a*. HPLC also facilitates the separation and quantification of up to 50 additional pigments in extracts of marine phytoplankton (Gibb *et al.*, 2000). The characterisation of these additional pigments allows for the distribution and composition of phytoplankton assemblages to be determined. For example 19' - hexanoyloxyfucoxanthin (HEX), fucoxanthin (FUC) and butanoyloxyfucoxanthin (BUT) have been used as biomarkers for prymnesiophytes, diatoms and chrysophytes respectively (Table 1.1) (Gibb *et al.*, 2000).



**Figure 1.6.** Schematic view of the proposed endosymbiotic events which have given rise to the diversity of plastids (taken from Keeling, 2010). Endosymbiotic events are depicted by being enclosed in boxes.

**Table 1.1. Chlorophyll and carotenoid chemotaxonomy of eukaryotic phytoplankton (adapted from Gibb *et al.*, 2000). S = major photosynthetic pigment, s = minor photosynthetic pigment, P = major photoprotectant pigment.**

Pigment	Diatoms	Prymnesiophytes	Prasinophytes	Dinoflagellates	Chrysophytes
chlorophyll $c_3$		S			S
chlorophyll $c_2$	S	S		S	S
chlorophyll $c_1$	S				
chlorophyll $b$			S		
chlorophyll $a$	S	S	S	S	S
Peridinin				S	
19'-Butanoyloxy-fucoxanthin		S			S
Fucoxanthin	S	S			s
19'-Hexanoyloxy-fucoxanthin		S			
Violaxanthin			S		
Prasinoxanthin			S		
Diadinoxanthin	P	P		P	P
Alloxanthin					
Diatoxanthin	P	P		P	
Zeaxanthin					P
Lutein			s		
$\beta$ catotene			P		

HPLC pigment signatures are often indicative of the class but several classes may share the same suite of pigments (e.g. diatoms and bolidophytes, Guillou *et al.*, 1999) or a given class may contain several pigment signatures (e.g. prasinophytes, Latasa *et al.*, 2004). Size fractionation studies, combined with HPLC pigment analysis, have allowed the composition of phytoplankton groups in different oceanic areas to be determined, based on algorithms such as CHEMTAX (Mackey *et al.*, 1996). Studies using HPLC have been used to analyse PPE diversity in coastal environments (Not *et al.*, 2007), the Mediterranean Sea (Brunet *et al.*, 2006), the Antarctic (Wright *et al.*, 2009) and the Atlantic Ocean (Poulton *et al.*, 2006). HPLC

pigment analysis has shown that prymnesiophytes were significant contributors to primary production in open ocean regions and especially in oligotrophic waters (Not *et al.*, 2008).

The use of microscopy and pigment studies can only assign broad taxonomic affiliations to organisms. In recent years, the development of molecular methods has enabled organisms to be classified to finer scales. One of the most popular markers to investigate the diversity of eukaryotes is the nuclear encoded small subunit (SSU) rRNA gene. This gene allows discrimination of diversity at a variety of taxonomic levels, due to the presence of both well conserved and rapidly evolving regions (Vaulot *et al.*, 2008). Another advantage of the use of the rRNA gene is that its transcribed product is a building block of ribosomes and present in high numbers, even in small cells.

Molecular investigation of marine eukaryotes only dates back to around 1998. As a by-product of bacterial (16S) rRNA gene studies several plastid-derived genes were found (Rappé *et al.*, 1998). The latter study was able to identify a relatively wide range of photosynthetic lineages related to the classes Bacillariophyceae, Cryptophyceae, Prymnesiophyceae, Chrysophyceae and Prasinophyceae. The first studies to specifically target eukaryotic organisms were undertaken to investigate the distribution of Bolidophyceae (Guillou *et al.*, 1999) and Prymnesiophyceae (Moon-van der Staay *et al.*, 2000), using general nuclear 18S rRNA gene primers and group-specific probes. Moon-van der Staay *et al.*, (2000) identified new clades of prymnesiophytes, with weak relationships to coccolithophorids and *Phaeocystis* which were not present in cultures, though generally the number of prymnesiophyte sequences found was low. This paucity of haptophyte sequences in SSU rRNA gene

clone libraries is in contrast to the abundance of their diagnostic pigment (19' HEX) in marine waters (Liu *et al.*, 2009). It was only recently, using haptophyte-specific nuclear LSU rRNA gene primers, that Liu *et al.*, (2009) showed a tremendous diversity of marine Haptophyta, with none of the environmental sequences identical to any of the taxonomically defined i.e. cultured, sequences. Most of the sequences formed new clusters related to *Chrysochromulina* clade B2 (Prymnesiales). The lineages Phaeocystales and Calcihaptophycidae also emerged in the picophytoplankton (Liu *et al.*, 2009).

More general use of universal 18S rRNA gene primers in the PCR has identified an unsuspected range of eukaryotic diversity, including phototrophic lineages other than prymnesiophytes, such as prasinophytes, stramenopiles and dinoflagellates as well as several heterotrophic taxa (Moon-van der Staay *et al.*, 2001). In studies using the nuclear 18S rRNA gene in coastal waters, the phototrophic component was dominated by sequences related to Chlorophyceae (see for example, Romari and Vaultot, 2004; Not *et al.*, 2007). Further studies, using samples obtained in the Antarctic and Atlantic, again showed Chlorophyceae to be the dominant class, but with members of the Prymnesiophyceae, Bacillariophyceae and dinoflagellates also significant contributors to the picophytoplankton (Diez *et al.*, 2001). Unfortunately, large proportions of nuclear 18S rRNA gene clone libraries are composed of heterotrophic sequences, such as novel marine stramenopiles (MAST) (Massana *et al.*, 2004) and marine parasitoids belonging to the Syndiniales (Alveolata) (Guillou *et al.*, 2008).

The dominance of heterotrophic sequences in nuclear 18S rRNA clone libraries has limited the ability to investigate photosynthetic lineages (Marie *et al.*, 2010). Indeed,

the abundance of heterotrophic sequences in environmental libraries is in contrast to microscopy counts which suggest that autotrophic cells account for, on average, 85% of the eukaryotic cells (Not *et al.*, 2004). One development to better assess the diversity of PPEs, is to combine the use of flow cytometric sorting with nuclear 18S rRNA clone library approaches. The sorting of PPE cells, using endogenous chlorophyll *a* as a basis of discrimination, saw the contribution of phototrophs in the clone libraries increase from 54% (from filtered samples) to 92% (Marie *et al.*, 2010). The reduction of heterotrophs in flow cytometry sorted samples also reduced the sampling effort required to identify photosynthetic sequences and this was highlighted with the recovery of several novel sequences in the sorted samples, which were not observed in filtered samples (Marie *et al.*, 2010; Shi *et al.*, 2009).

Since the first identification of algal-derived plastid 16S rRNA gene sequences, as a by-product of general bacterial studies, it was only recently that Fuller *et al.*, (2006a) designed primers that were biased towards amplifying the plastids of marine algae. These primers were used to target PPEs in the Arabian Sea (Fuller *et al.*, 2006a). In this study chrysophytes were a significant component of the PPE community. Previous 18S rRNA studies had indicated the presence of only heterotrophic forms of chrysophytes (e.g. related to *Paraphysomonas*) (Diez *et al.*, 2001) and only a few studies had identified photosynthetic forms, related to the order Parmales (Bravo-Sierra and Hernandez-Becceril, 2003). Further studies, in a variety of environments, have been undertaken using these plastid biased 16S rRNA gene primers. Prymnesiophyceae and Chrysophyceae dominate these libraries in the Mediterranean Sea (McDonald *et al.*, 2007), Atlantic (Kirkham *et al.*, 2011) and Pacific Oceans (Shi *et al.*, 2011). Other lineages which contributed to the clone libraries, include



Pelagophyceae, Dictyochophyceae and Prasinophyceae. Many of the sequences obtained in these libraries have no close relatives in described species, with several novel lineages identified e.g. the prasinophyte clade 16S-IX (Shi *et al.*, 2011).

An alternative molecular method to target organisms in the marine environment has been to target the *psbA* gene, encoding the D1 polypeptide of photosystem II (Zeidner *et al.*, 2003). These authors found *psbA* sequences derived from both prokaryotic and eukaryotic organisms in their clone libraries. The eukaryotic component comprised mainly Prasinophyceae, although sequences relating to Prymnesiophyceae, Cryptophyceae and Stramenopiles were also observed. Prymnesiophyceae, in agreement with pigment studies, were found to make a large contribution to *psbA* transcript clone libraries in the Mediterranean Sea, with significant contributions from the classes Pelagophyceae and Chrysophyceae (Man-Aharonovich *et al.*, 2010).

Oligonucleotide probes have also been designed to target specific algal taxa and have been used to estimate the abundance of PPEs in various environments (Lim *et al.*, 1996; Fuller *et al.*, 2006b). Probes hybridise to PCR-amplified rDNA immobilised on membranes (dot-blots) or directly to rRNA from fixed whole cells (*in situ* hybridisation) (Biegala *et al.*, 2003). Fluorescence *in situ* hybridisation (FISH) involves the direct labelling of the probe with a fluorochrome. The fluorescent signal can be amplified by the deposition of tyramide bound fluorochromes, in the proximity of horseradish peroxidase labelled probes, to improve the detection of small cells (Schönhuber *et al.*, 1997). Simon *et al.*, (1995) were able to distinguish between chlorophyta and non chlorophyta taxa in combination with flow cytometry. Moreover, dot blot hybridisation analysis of PPEs along an Atlantic Meridional

Transsect (AMT) transect showed an interesting complimentary distribution between the dominant classes Prymnesiophyceae and Chrysophyceae, with Chrysophyceae dominating northern temperate waters, the southern gyre and equatorial regions, whilst prymnesiophytes dominated the northern gyre (Kirkham *et al.*, 2011). Multivariate statistical approaches have subsequently given insights into the environmental factors potentially affecting the distribution of specific PPE groups. In the Atlantic Ocean Chrysophyceae were found to correlate with higher light levels, with a decrease in abundance with depth (Kirkham *et al.*, 2011) whilst in the Pacific Ocean, Chrysophyceae was linked to dissolved oxygen concentration (Lepère *et al.*, 2009). In both studies, the majority of the variation in PPE distribution patterns was unexplained by the factors measured, and no measured factors were able to explain the distribution of Prymnesiophyceae. The amount of unexplained variation may also be due the high genetic diversity present within each class, potentially equating to taxa with dissimilar physiologies that respond differently to environmental variables, as well as unmeasured factors such as predation, viral lysis and competition (Lepère *et al.*, 2009).

### **1.8. Genomics of marine phytoplankton**

Recently, there has been increased interest in the genomics of marine phytoplankton and sequencing of several genomes of the cyanobacteria *Synechococcus* (see Palenik *et al.*, 2003; Palenik *et al.*, 2006, Dufresne *et al.*, 2008) and *Prochlorococcus* (Rocap *et al.*, 2003; Kettler *et al.*, 2007) have been undertaken. The availability of a variety of genomes representing clades adapted to different environments, enabled a comparative genomic approach to be undertaken identifying core and accessory genes (Dufresne *et al.*, 2008). A comparison of 11 *Synechococcus* and three

*Prochlorococcus* genomes revealed a core set of genes which were present in all 14 of the cyanobacterial genomes analysed. The core picocyanobacterial genome consisted of 1,228 gene families. In addition 70 gene families were found in all the *Synechococcus* genomes but absent from the *Prochlorococcus* genomes including 23 linked to photosynthesis (Dufresne *et al.*, 2008). The remainder of the genome was comprised of accessory genes (found in 2-10 strains) and unique genes (found in only one strain), suggesting these genes are probably relevant in the adaptation to specific environments (Dufresne *et al.*, 2008). Comparative genomics was able to show, for example, that the coastal *Synechococcus* strain CC9311 had a higher number of metal enzymes, cofactors and transporters than the open ocean strain WH8102 (Palenik *et al.*, 2006).

Algal genomics began with the sequencing of the 551-kilobase remnant nucleomorph genome of the cryptomonad *Guillardia theta* (Douglas *et al.*, 2001). Since then, only 12 genomes of eukaryotic algae have been completely sequenced, including nine marine and two freshwater species (Table 1.2).

**Table 1.2. Species of algae which have had their genomes published**

Species	Reference
<i>Aureococcus anophagefferens</i>	Gobler <i>et al.</i> , 2011
<i>Micromonas pusilla</i> (CCMP1545)	Worden <i>et al.</i> , 2009
<i>Thalassiosira pseudonana</i>	Armbrust <i>et al.</i> , 2004
<i>Phaeodactylum tricornutum</i>	Bowler <i>et al.</i> , 2008
<i>Ostreococcus lucimarinus</i>	Derelle <i>et al.</i> , 2006
<i>Volvox carteri</i>	Prochnik <i>et al.</i> , 2010
<i>Emiliana huxleyi</i>	Worden and Allen, 2010
<i>Ostreococcus</i> sp.	Palenik <i>et al.</i> , 2007
<i>Chlamydomonas reinhardtii</i>	Merchant <i>et al.</i> , 2007
<i>Micromonas pusilla</i> (RCC299)	Worden <i>et al.</i> , 2009
<i>Cyanidioschyzon merolae</i>	Matsuzaki <i>et al.</i> , 2004
<i>Chlorella</i> sp. NC64A	Blanc <i>et al.</i> , 2010

Among the microalgae which have had their genomes sequenced are two diatom species, *Thalassiosira pseudonana* (Armbrust *et al.*, 2004) and *Phaeodactylum tricornutum* (Bowler *et al.*, 2008). *T. pseudonana* has a genome size of 32.4 Mb with 11,776 predicted genes, while *P. tricornutum* has a slightly smaller genome size of 27.4 Mb with 10,402 predicted genes (Bowler *et al.*, 2008). About 57% of the *P. tricornutum* genes are shared with *T. pseudonana*, with 1,328 being diatom specific (Bowler *et al.*, 2008). One of the major findings from the genomes of the two diatom species is the presence of a urea cycle which had not been previously found in a phototrophic eukaryote (Armbrust *et al.*, 2004). It has been suggested that the presence of a urea cycle could recover  $\text{NH}_3$  and  $\text{CO}_2$  mitochondrial photorespiratory products, through the activity of carbamoyl phosphate synthase or carbamate kinase (Allen *et al.*, 2006). Furthermore, the sequencing of the genomes identified putative  $\text{C}_4$  metabolism in diatoms, where decarboxylation of oxaloacetate (OAA) and malate occurs in the mitochondria and  $\text{CO}_2$  utilization by RuBisCO occurs in the plastid (Kroth *et al.*, 2008).

Expression profiling of *T. pseudonana* was used to identify the genes involved in silicon bioprocesses (Mock *et al.*, 2008). 75 genes were up-regulated under silicon limitation but with the majority of these genes without predicted function. The authors were also able to identify 84 genes which were up-regulated under both silicon and iron limitation but under no other treatment, suggesting a tight coupling between pathways initiated by iron and silicon bioavailability (Mock *et al.*, 2008).

Five of the currently sequenced algal genomes relate to species which are classified as PPEs. These are the Mamiellophyceae genomes *Ostreococcus lucimarinus*,

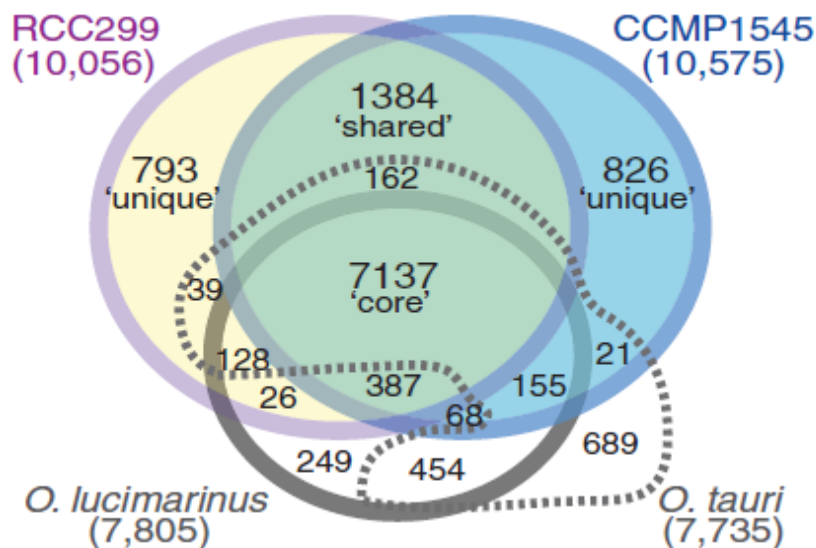
*Ostreococcus tauri*, and two strains of *Micromonas pusilla* (RCC299 and CCMP1545) and the Pelagophyceae genome *Aureococcus anophagefferens*.

*Ostreococcus tauri* is the smallest free living eukaryote measuring <1  $\mu\text{m}$  in diameter (Courties *et al.*, 1994) and has a genome size of only 12.6 Mb (Derelle *et al.*, 2006), with *Ostreococcus lucimarinus* having a slightly larger nuclear genome of 13.2Mb (Palenik *et al.*, 2007). The genomes are predicted to have 7,892 and 7,651 nuclear genes, respectively (Peers and Niyogi, 2008), which is considerably more than the oxygenic prokaryote *Prochlorococcus* (1.7Mb and ~1900 predicted ORFs, Dufresene *et al.*, 2003). The sequencing of the *Ostreococcus* genome showed the intense degree of genome compaction present in this organism. Several processes, such as the shortening of intergenic regions, gene fusion and the reduction of the size of gene families, are believed to be important in this compaction (Derelle *et al.*, 2006). Comparisons between the two *Ostreococcus* genomes showed that 5-6% of the predicted genes did not show homology to the other species, despite having 99.8% identity in 18S rRNA gene sequences (Palenik *et al.*, 2007). One of the interesting findings from the *Ostreococcus* genomes was the presence of the machinery necessary to perform  $\text{C}_4$  photosynthesis, including a putative cytosolic phosphoenolpyruvate carboxylase and at least one chloroplast targeted NADP-dependent malic enzyme (Archibald, 2006).

*Micromonas* isolates RCC299 and CCMP1545 are from geographically distant oceanic provinces and were generally considered to be a single species, based on 18S rRNA gene studies where they had 97% identity (Worden *et al.*, 2009). However, on sequencing the two genomes, it was found that only 90% of their 10,056 (RCC299) and 10,575 (CCMP1545) predicted genes were shared. The divergence between the

predicted genes supports the idea that these two isolates should be classified as separate species (Worden *et al.*, 2009). Phylogenetic analysis of genes which only occurred in one of the strains, showed that they were distantly related to organisms such as animals, fungi and bacteria. One interpretation of this finding is that the genes are a product of horizontal gene transfer (Archibald, 2009b).

Comparative analysis of the four Mamiellophyceae genomes (2 *Ostreococcus* and 2 *Micromonas*) showed that there was a core genome of 7,137 genes (Figure 1.7) (Worden *et al.*, 2009).



**Figure 1.7. Venn diagram showing a comparison of the gene compliments of four sequenced Mamiellophyceae genomes (from Worden *et al.*, 2009).**

The two *Micromonas* genomes shared 1,384 (14%) genes which were not shared with the *Ostreococcus* genomes (Worden *et al.*, 2009). Compared with *Ostreococcus*, *Micromonas* contains a more complex set of nutrient transporters, as well as more genes relating to combating reactive oxygen species and heavy metals (Archibald, 2009b). *Micromonas* has a broader global distribution than *Ostreococcus*

and the increased number of genes may make it more flexible in terms of environmental adaptability (Archibald, 2009b).

Although *Aureococcus anophagefferens* is a similar size ( $\sim 2 \mu\text{m}$ ) to the four Mamiellophyceae species, the genome of *A. anophagefferens* is considerably larger (56 Mbp) and has more genes ( $\sim 11,500$ ) predicted in its genome (Gobler *et al.*, 2011). *A. anophagefferens* is of particular interest due to its ability to produce harmful algal blooms (HABs) in estuarine regions. The genome revealed a suite of 62 light harvesting complex genes which is 1.5-3 times more than other eukaryotic phytoplankton and may be an adaptation to the low light levels experienced during bloom conditions (Gobler *et al.*, 2011). The *A. anophagefferens* genome also contains a collection of genes enabling it to outcompete other phytoplankton species in regions with elevated organic matter. Thus, the genome has a set of transporters for organic nitrogen including urea, amino acids, purines and oligopeptides (Gobler *et al.*, 2011). Relative to other phytoplankton, the genome is also enriched in enzymes which degrade organic nitrogen such as urea, amino acids and proteins (Gobler *et al.*, 2011). In estuarine environments, concentrations of trace elements such as selenium are elevated (Cutter and Cutter, 2004) and *A. anophagefferens* has a higher selenium requirement than other phytoplankton (Gobler *et al.*, 2011). This is reflected in the genome by the enrichment of proteins which require selenium as a cofactor including several selenoproteins which have never before been described (Gobler *et al.*, 2011). Selenocysteine residues are often catalytically superior compared to cysteine (Stadtman, 1996), and in regions of elevated selenium the production of these superior enzymes may be an important reason for the formation of HABs.

Genomic analysis of cultured phytoplankton allows the specific annotation of gene function. Currently, the dearth of available genomes limits the ability to assign functions to many genes. In eukaryotes, about 30% of genes have no known function with a further 10% poorly characterised (Worden and Allen, 2010). However, comparative studies have been able to create hypotheses about the function of specific genes. In the diatom *Thalassiosira pseudonana*, transcriptomic studies identified unexpected links between iron and silicon limitation and identified several novel silicon bioprocessing components (Mock *et al.*, 2008). The functional annotation of genes from cultured representatives also increases the ability to annotate reads obtained from metagenomic and metatranscriptomic studies. Projects such as Tara Oceans (<http://oceans.taraexpeditions.org/>) may in the future increase the number of genomes from marine phytoplankton which would then allow for improved analysis of the functional capability of environmental systems.

### **1.9 History of microbial environmental metagenomics**

With the development of high-throughput sequencing and advances in DNA amplification technologies, metagenomics, the genomic analysis of a population of microorganisms, has emerged as a powerful tool (Handelsman, 2004). Uncharacterised microbial communities contain the vast majority of the biosphere's genetic material and metagenomics is enabling the composition and dynamics of the predominantly unculturable microbial community to be deciphered (Deutschbauer *et al.*, 2006). The seminal study in metagenomics was undertaken by Tyson *et al.*, (2004) on an acid mine drainage (AMD) site. The focus on a low complexity AMD site enabled the assembly of an almost complete genome of two species, including the dominant species (a *Leptospirillum* group II species) (Tyson *et al.*, 2004).



Nitrogen fixation is a process essential to the AMD community. Interestingly the dominant group (*Leptospirillum* group II) did not contain any genes responsible for nitrogen fixation. A complete operon for nitrogen fixation was however, observed in sequences related to the relatively low abundance group *Leptospirillum* group III (Tyson *et al.*, 2004). Since the first metagenomic study more complicated environments have been investigated, including the Sargasso Sea (Venter *et al.*, 2004) and a comparison made between the microbial communities in agricultural soil and deep sea whale skeleton environments (Tringe *et al.*, 2005). The more complex environments did not result in assembled genomes but they did allow the identification of gene families important for overall survival in the environment (Tringe and Rubin, 2005). The information generated from metagenomic projects in the four years preceding 2009, led to the identification of more than  $40 \times 10^6$  genes, far exceeding the  $9 \times 10^5$  genes resulting from genome sequencing or traditional culturing methods (Guazzaroni *et al.*, 2009). Metagenomic studies, from a wide range of environments have also been used to probe for novel drugs and enzymes with potential uses in the biotechnological field (Schmeisser *et al.*, 2007).

The next step in the meta-omics approach is the analysis of community transcripts (metatranscriptomics). This method moves beyond investigating the genomic potential of the community and allows for the correlation of the *in situ* activity (function) with the environmental conditions of a specific environment (Chistoserdova, 2010). The first study to employ pyrosequencing for a metatranscriptomic study was Leininger *et al.*, (2006), where they identified the activity and importance of ammonia oxidising archaea. The natural abundance of rRNA enables a detailed phylogenetic profiling of the sample (e.g. see Urich *et al.*,

2008). However, to gain a functional analysis of the sample often requires the amplification of mRNA (e.g. see Frias-Lopez *et al.*, 2008) and typically only a third of the potential mRNA transcripts can be matched to known genes or gene families (Chistoserdova, 2010). Recently, metatranscriptomic studies have been used to study specific metabolic pathways in the environment. McCarren *et al.*, (2010) investigated the microbes and metabolic pathways associated with dissolved organic matter (DOM). Increases in phosphate two component systems and phosphate ABC transporters were observed, with enrichment of samples with DOM (McCarren *et al.*, 2010). One of the first metatranscriptomic studies investigating eukaryotic communities was undertaken by Grant *et al.*, (2006) on geothermal algal mats and activated sludge. mRNA was enriched in the samples taking advantage of the poly (A) tail of eukaryotic mRNAs, enabling a better investigation of the functional properties of the sample (Grant *et al.*, 2006).

Metaproteomics, the analysis of community protein profiles, developed as the next approach to investigating the immediate catalytic potential of the community (Simon and Daniel, 2011). One of the most comprehensive metaproteomic studies was also undertaken on an AMD site (Ram *et al.*, 2005). Of the abundant proteins observed in the metaproteomics sample, 15% of them had no similarity to known proteins. Metaproteomic studies of more complex environments included an analysis of the microbial community in Chesapeake Bay (Kan *et al.*, 2005). The Chesapeake Bay study was able to identify differences in the protein composition at three different stations in the bay and highlighted the future possibilities of inferring functional characteristics of a station (Kan *et al.*, 2005). A further study by Sowell *et al.*, (2009) in the Sargasso Sea showed a high abundance of proteins involved in photosynthesis

and carbon fixation which were related to the cyanobacteria *Prochlorococcus* and *Synechococcus*.

Further development of the complete set of “omics” technologies, alongside studies of cultured representatives, will allow the genetic and functional properties of microbial communities to be linked, and to relate taxonomic and functional diversity to ecosystem stability (Schneider and Riedel, 2010).

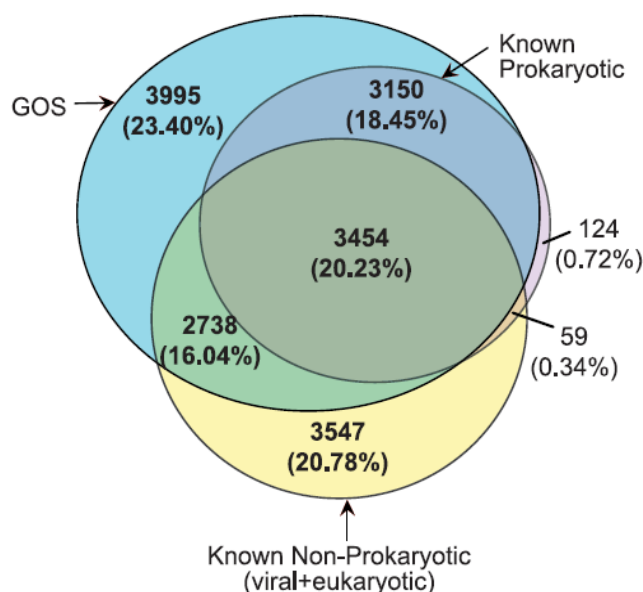
### **1.10 Metagenomics of marine phytoplankton**

The majority of microorganisms are only known from molecular studies and no representative culture of many environmental groups are available. Direct isolation of environmental genomic DNA allows for the circumvention of culturing in the study organisms. With an appropriate depth of sequencing, metagenomic studies can lead to the construction of genomes from environmentally important organisms. Metagenomic studies are able to identify the functional potential of the organisms in the environment and can help identify which organisms are undertaking functional roles in the environment.

Shotgun sequencing of the microbial fraction of the Sargasso Sea was one of the earliest metagenomic studies to encompass phytoplankton (Venter *et al.*, 2004). Approximately 1.6 million reads were obtained from four stations in the Sargasso Sea, which when pooled generated ~ 64,000 scaffolds, ranging from 826 bp to 2.1 Mbp in length. Scaffolds relating to several species were identified, including *Burkholderia*, *Shewanella*, SAR86 and a conglomerate of *Prochlorococcus* strains (Venter *et al.*, 2004), although the *Burkholderia* and *Shewanella* sequences may be related to ship-board contamination (DeLong and Karl, 2005). Venter *et al.*, (2004)

focussed on the prokaryotic component of the sample by filtration through a 0.8  $\mu\text{m}$  filter. However, Piganeau *et al.*, (2008) analysed the Sargasso Sea Database (SSD) to assess the diversity and abundance of eukaryotes in the sample. Genomic data from at least 41 eukaryotes were found in the 0.8  $\mu\text{m}$  filter fraction, with representatives from the five supergroups of the eukaryote tree: Rhizaria, Chromalveolates, Excavates, Plantae and Unikonts (Piganeau *et al.*, 2008). The 8,166 predicted protein coding genes in the *Ostreococcus* genome were analysed by BLAST against the SSD, with 23% of the genome and 14% of the predicted proteins having a match in the SSD (Piganeau and Moreau, 2007). Overlapping chloroplast contigs enabled the identification of two different *Ostreococcus* strains to be present in the sample (Piganeau and Moreau, 2007). The Sargasso Sea sampling of Venter *et al.*, (2004) was a pilot project for the ambitious Sorcerer II Global Ocean Sampling (GOS) expedition (Rusch *et al.*, 2007). The aim of the GOS expedition was to sample all the global ocean basins and the first stage studied a several thousand km long transect from the North Atlantic, through the Panama Canal and into the South Pacific (Rusch *et al.*, 2007). Approximately 7.7 million sequencing reads were obtained from the transect, with 85% of the assembled sequences being found to be unique at the 98% sequence identity cut-off (Rusch *et al.*, 2007). This large sequencing effort greatly expanded the number of known protein families (Yooseph *et al.*, 2007). Comparing the medium and large clusters (size >20) of the GOS database against the known eukaryotic and prokaryotic databases, showed that 3,995 protein clusters were unique to the GOS database (Figure 1.8). The large number of protein clusters, which could not be assigned to protein families, highlights the potential diversity still to be found and has added considerable knowledge to existing protein families (Yooseph *et al.*, 2007). For example, there are forms of RuBisCO in the GOS

database which do not cluster with known proteins of this family. Thus the RuBisCO protein family may be more diverse than currently identified (Yooseph *et al.*, 2007).



**Figure 1.8. Breakdown of the 17,067 medium and large clusters from the GOS database into three categories – GOS, known prokaryotic and known eukaryotic (from Yooseph *et al.*, 2007). Percentages for each group are given in brackets.**

Recently Cuvelier *et al.*, (2010) undertook a targeted metagenomic approach to understanding the functional potential of eukaryotic picophytoplankton. Pico-prymnesiophytes were targeted using flow cytometry with the majority of rRNA sequences being related to uncultured prymnesiophytes (Cuvelier *et al.*, 2010). Some of the 1,624 predicted proteins indicated adaptations to life in oceanic environments. Superoxide dismutases (SODs) are important for scavenging toxic superoxide radicals and different isoforms use a variety of metals. In the prymnesiophyte metagenome only Ni-SODs were observed, with a distinct absence of Fe-SODs (Cuvelier *et al.*, 2010). This may result in the reduction in the cellular demands for

Fe, which is in lower concentration in the open ocean compared with Ni (Dupont *et al.*, 2010).

### 1.11 Metatranscriptomics of marine phytoplankton

Metagenomic studies have been able to reveal some of the functional potential of microorganisms in the environment. However, it was not until the development of metatranscriptomic approaches that there was an understanding of the expression of genes in the environment. The functional capability of organisms in the environment allows for analysis of what is driving the distribution of organisms in the marine environment.

Frias-Lopez *et al.*, (2008) undertook an analysis of a metatranscriptome from surface waters of the North Pacific subtropical gyre. Only 13% of the reads could be matched to the NCBI nr database, while 43% were matched against the GOS database, which may reflect the presence of rare or novel genes (Frias-Lopez *et al.*, 2008). The taxonomic origins of those reads which could be assigned were related to *Prochlorococcus* and the  $\alpha$ -proteobacteria *Pelagibacter*. The majority of the *Prochlorococcus* sequences were related to the high light ecotype, in agreement with depth profile ecotype abundance data (Johnson *et al.*, 2006). Comparison of the metatranscriptome against the core *Prochlorococcus* genome showed that transcripts involved in photosynthesis and carbon fixation were highly represented in the library (Frias-Lopez *et al.*, 2008).

Poretsky *et al.*, (2009) undertook a day/night metatranscriptome comparison in the North Pacific gyre. Of the 167 KEGG pathways which were represented in the samples only four were better represented at night, including glycosphingolipid

biosynthesis and nucleotide sugar metabolism. In the daylight sample, six pathways had increased representation including photosynthesis and oxidative phosphorylation. Overall, this oligotrophic system was skewed towards energy metabolism during the day and biosynthesis at night (Poretsky *et al.*, 2009).

Analysis of eukaryotic metatranscriptomes is lagging behind their prokaryotic counterparts. A pilot project for the creation of a eukaryotic metatranscriptome in a eutrophic bay (Tampa Bay, Florida) was undertaken by John *et al.*, (2009). Only about 25.7% of the reads had matches against the NCBI nr database. Of those sequences which were assigned, 70% were phylogenetically related to diatoms, with smaller numbers related to dinoflagellates and chlorophytes (John *et al.*, 2009). Functional characterisations of the reads, showed that protein synthesis/modification/degradation were the most frequent category present in the sample. This small scale pilot study has set the ground work for larger scale efforts into investigating the transcriptomic profile of eukaryotic phytoplankton. Increased knowledge of the functional aspects of eukaryotic phytoplankton will lead to an improved understanding of biogeochemical cycles in the marine environment and especially the carbon cycle where PPEs are such a significant component.

### **1.12 Overall aims**

The main objective of this study was to elucidate the taxonomic composition and distribution of specific PPE groups over large spatial scales so as to begin to understand potential factors important in dictating their presence/absence in specific environments, and thus help to explain controls on marine photosynthesis. In so doing, we aimed to begin to identify the specific roles of different members of the PPE community in global fluxes of carbon and nutrients. As well as using standard

molecular ecological approaches this project entailed the development of new molecular pipelines to further characterise the physiology of PPEs in natural oceanic environments, as well as attempts to obtain novel, but environmentally relevant PPEs, into enrichment culture.

#### 1.12.1 Specific aims

- 1) To obtain taxonomic information on specific flow-sorted populations from the SOLAS cruise in the sub-tropical North Atlantic in 2008 so as to couple phylogenetic identification of specific PPE populations with  $^{14}\text{C}$ -radiotracer uptake measurements into these groups i.e. to directly link  $\text{CO}_2$  fixation with taxonomy (Chapter 3).
- 2) To construct marine-algal plastid biased 16S rRNA and nuclear 18S rRNA gene clone libraries on specific flow sorted PPE populations along AMT18 in the Atlantic Ocean in order to establish the composition and distribution of specific PPE groups at the ocean basin-scale (Chapter 4).
- 3) To correlate these distribution patterns with a range of environmental factors so as to begin to ascertain the factors driving the distribution of these organisms in the environment (Chapter 4).
- 4) To attempt to isolate into culture novel strains/genera of PPEs using enrichments from environmental samples along AMT18, for future physiological characterisation (Chapter 5).
- 5) To undertake a transcriptomic approach to analysing two PPE cultures potentially representative of groups widespread and/or abundant along AMT18. The approach



will provide a molecular means to inform us of the potential physiology of these organisms as well as to inform future environmental metatranscriptomic studies by allowing phylogenetic assignment of sequence reads (Chapter 6).

6) To develop an environmental metatranscriptomic pipeline specifically targeting PPEs from AMT18 and to use subsequent metatranscriptomic information for assessing ‘the environmental variables these organisms are responding to’ (Chapter 7).

## **Chapter 2**

### **Methods**

## 2.1 Culturing

### 2.1.1 Growth and maintenance of *E. coli* strains

The genotype of the three *E. coli* strains used in this thesis is described in Table 2.1.

Luria Broth (LB) was used for the routine growth and maintenance of *E. coli* strains.

The composition of LB medium is 10 g L<sup>-1</sup> tryptone, 5 g L<sup>-1</sup> yeast extract and 10 g L<sup>-1</sup> NaCl. The pH was adjusted to 7.0 before sterilisation by autoclaving at 121°C for 15 min. The LB medium was supplemented with antibiotics, as described, when applicable.

SOC medium was used for recovery growth of *E. coli* strains after electroporation in the clone library experiments. The composition of SOC is 20 g L<sup>-1</sup> tryptone, 5 g L<sup>-1</sup> yeast extract, 0.5 g L<sup>-1</sup> NaCl, 0.1875 g L<sup>-1</sup> KCl, 1 g L<sup>-1</sup> MgCl<sub>2</sub>, 1.2 g L<sup>-1</sup> MgSO<sub>4</sub> and 3.6 g L<sup>-1</sup> glucose.

**Table 2.1. Genotype of the three *E. coli* strains Top10F', JM109 and DH10B T1 Phage Resistant used in this study.**

Strain	Genotype	Supplier	
Top10F'	F' { <i>lacIq</i> Tn10 (Tet <sup>R</sup> ) } <i>mcrA</i> Δ( <i>mrr-hsdRMS-mcrBC</i> ) Φ80 <i>lacZ</i> ΔM15 Δ <i>lacX74</i> <i>recA1</i> <i>araD139</i> Δ( <i>ara-leu</i> )7697 <i>galU</i> <i>galK</i> <i>rpsL</i> <i>endA1</i> <i>nupG</i>	Invitrogen	SOLAS clone library
JM109	<i>recA1</i> , <i>endA1</i> , <i>gyrA96</i> , <i>thi</i> , <i>hsdR17</i> (rK <sup>-</sup> ,mK <sup>+</sup> ), <i>relA1</i> , <i>supE44</i> , Δ( <i>lac-proAB</i> ), [F', <i>traD36</i> , <i>proAB</i> , <i>lacIqZ</i> ΔM15]	Promega	AMT clone library
DH10B T1 Phage Resistant	F- <i>mcrA</i> Δ( <i>mrr-hsdRMS-mcrBC</i> ) Φ80 <i>lacZ</i> ΔM15 Δ <i>lacX74</i> <i>recA1</i> <i>endA1</i> <i>ara</i> Δ139 Δ( <i>ara, leu</i> )7697 <i>galU</i> <i>galK</i> λ- <i>rpsL</i> <i>nupG</i> <i>tonA</i>	Invitrogen	cDNA library construction

### **2.1.2 Isolation of photosynthetic picoeukaryote (PPE) cultures**

#### 2.1.2.1 Collection of seawater samples for culturing during the AMT18 cruise

During the AMT18 cruise (October–November 2008), water samples for cultures were taken from the 55% light level and the deep chlorophyll maximum (DCM) during a pre-dawn CTD, using 20L Niskin bottles. The 55% light level was calculated from the previous day's noon CTD cast. Cells were concentrated, whilst on board, using two different methods:

#### 2.1.2.2 PPE cell sorting using flow cytometry

PPEs were isolated from seawater samples using a FacSort flow cytometer (Becton Dickinson, US) at various stations during AMT 18. Using endogenous chlorophyll fluorescence (FL3) and side scatter (SSC) to discriminate the photosynthetic picoeukaryotes, between 5000 and 50000 cells were sorted into capped glass test tubes. K artificial seawater medium (see section 2.1.2.4) was added to each tube. Vials were then incubated as described below. All isolation tubes were immediately incubated at 21°C, under a constant light intensity of  $30 \mu\text{mol m}^{-2} \text{s}^{-1}$ .

#### 2.1.2.3 Concentration from natural seawater samples

In order to target the pico size fraction of the plankton, cells from pre-filtered seawater samples (through 10  $\mu\text{m}$  and 5  $\mu\text{m}$  pore size polycarbonate filters (PALL)) were further filtered onto a 0.22  $\mu\text{m}$  Supor filter (PALL). Cells were washed off the filter using 2 ml of 0.22  $\mu\text{m}$  filtered deep water and then diluted, using K medium, into plastic 50 ml culture flasks with 0.22  $\mu\text{m}$  filter lids (BD Biosciences). All isolation flasks were immediately incubated at 21°C, under a constant light intensity of  $30 \mu\text{mol m}^{-2} \text{s}^{-1}$ .

#### 2.1.2.4 Growth of PPEs

To enhance the growth of different PPEs in enrichment cultures, sub cultures were established into different media. The composition of the media used is described in Table 2.2. Nutrients were added to artificial seawater (Sigma) and autoclaved for 15 min at 121°C. The media were allowed to cool before filter sterilised trace metals and vitamins were added. In the case of the soil extract medium (Andersen, 2005) the soil extract was added after autoclaving.

#### **2.1.3 Maintenance of PPE cultures**

Routinely PPE cultures were maintained at 21°C, under a light intensity which was set at 30  $\mu\text{mol m}^{-2} \text{s}^{-1}$  with a 12 h light/dark cycle. A replicate set of the culture collection was maintained in ambient laboratory conditions, at room temperature and under room lights.

#### **2.1.4 Purification of agar**

Commercial Bacto agar (Becton Dickinson) (250 g) was suspended in 5 L of MilliQ water and stirred for 30 min. The agar was allowed to settle and the water was poured off. The agar was then filtered through 3MM Whatman (Whatman) filter paper, using a Buchner funnel over a side arm flask with the aid of a vacuum pump. Repeated washes with water were undertaken until the flow through of water was clear. The cleaned agar was then stirred for 30 min in 5 L of ethanol before being filtered through 3MM Whatman paper and then further washed for 30 min in 5 L acetone. The washed agar was filtered through 3MM Whatman paper and left to dry completely in a fume hood. Cleaned agar was stored in an air tight container until required.

**Table 2.2. Composition of PPE growth media. Final concentrations of nutrients are given.**

Component	K Medium (Keller <i>et al</i> , 1987)	f/2 Medium (Guillard and Ryther, 1962)	PC Medium (Andersen, 2005)	Soil Extract Medium (Andersen, 2005)
Seawater (Sigma)	993.5ml	996.5ml	997ml	990ml
NaNO <sub>3</sub>	882μM	882μM		882μM
NaH <sub>2</sub> PO <sub>4</sub> ·H <sub>2</sub> O		36.2μM		36.2μM
NH <sub>4</sub> Cl	26.7μM		50μM	
Na <sub>2</sub> β-glycerophosphate	10μM		10μM	
H <sub>2</sub> SeO <sub>3</sub>	100nM		1nM	
Tris-base (pH7.2)	1mM			
Urea			50μM	
Soil water extract (see section 5.2.1)				10ml
FeCl <sub>3</sub> ·6H <sub>2</sub> O	11.7μM	11.7μM	11.7μM	
Na <sub>2</sub> EDTA·2H <sub>2</sub> O	11.7μM	11.7μM	112μM	
MnCl <sub>2</sub> ·4H <sub>2</sub> O	910nM	910nM	99nM	
ZnSO <sub>4</sub> ·7H <sub>2</sub> O	76.5nM	76.5nM	7.65nM	
CoCl <sub>2</sub> ·6H <sub>2</sub> O	42.0nM	42.0nM	4.20nM	
CuSO <sub>4</sub> ·5H <sub>2</sub> O	39.3nM	39.3nM		
Na <sub>2</sub> MoO <sub>4</sub> ·2H <sub>2</sub> O	26.0nM	26.0nM	2.6nM	
NiSO <sub>4</sub> ·6H <sub>2</sub> O			1nM	
Thiamine·HCl (vitamin B <sub>1</sub> )	296nM	296nM	29.6nM	
Biotin (vitamin H)	2.05nM	2.05nM	0.205nM	
Cyanocobalamin (vitamin B <sub>12</sub> )	0.369nM	0.369nM	0.0369nM	

### **2.1.5 Agar plating**

#### 2.1.5.1 Preparation of agar

Agar plates were prepared by supplementing liquid media (Table 2.2) with 9 g L<sup>-1</sup> of purified agar. The agar was sterilised by autoclaving for 15 min at 121°C and allowed to cool. Filter sterilised trace metals and vitamins (Table 2.2) were added separately to the cooled media. Cultures were plated using two different methods.

#### 2.1.5.2 Culturing of picoeukaryotes using pour plates

An aliquot of liquid culture was mixed gently in a Petri dish with warm medium (~40°C), supplemented with agar. Once the medium set, the plates were incubated at 21°C with a 12 h light/dark cycle under a light intensity of 30  $\mu\text{mol m}^{-2} \text{s}^{-1}$ . Cells were retrieved from the agar using a modified glass pasteur pipette. The pipette was flamed and stretched until a thin tube was obtained, in which a single colony could be isolated. The single colony was transferred into 50 ml plastic culture flasks containing 10 ml of the appropriate medium.

#### 2.1.5.3 Culturing of picoeukaryotes using solid agar.

Solid medium, supplemented with 0.9% (w/v) agar, was prepared in Petri dishes. Cultures were streaked onto the surface of the agar plate using sterile spreading loops. Plates were incubated at 21°C with a 12 h light/dark cycle under a light intensity of 30  $\mu\text{mol m}^{-2} \text{s}^{-1}$ . Single colonies were retrieved from the surface of the plate and placed into 50 ml culture flasks containing the appropriate medium.

## **2.2 Molecular Techniques**

### **2.2.1 DNA extraction protocols**

#### **2.2.1.1 Fast chemical denaturation protocol for rapid analysis of cultures**

An aliquot of culture (1.5ml) was centrifuged at 16060 x *g* in a Biofuge Pico (Heraeus) for 5 min after the addition of 1% (w/v) Pluronic (final concentration) to avoid cell adhesion on the surface of the tubes. The supernatant was discarded and the cells were resuspended in 5 µl of denaturation solution (400mM KOH, 10mM EDTA). The mix was incubated on ice and then 5 µl of neutralisation buffer (400mM HCl, 600mM Tris-HCl pH 9.0) was added. A 1/10 dilution in MilliQ H<sub>2</sub>O was then used as a template for PCR reactions.

#### **2.2.1.2 Nucleic acid extraction for sorted samples from the SOLAS cruise (2008)**

PPE cells sorted by flow cytometry (see section 3.2.1) were added to 50 µl of lysis buffer (20 mM EDTA, 400 mM NaCl, 0.75 M sucrose, 50 mM Tris-HCl pH 9.0 sterilised by autoclaving), 6 µl SDS (10% w/v), 10 µl Proteinase K (10 mg ml<sup>-1</sup>). The reaction mix was incubated at 37°C for 30 min and then a further 10 min at 55°C. Nucleic acids were precipitated at -20°C, for 10 min, with 2 µl glycogen (20 mg ml<sup>-1</sup>) and 2.5 vol. of absolute ethanol. The nucleic acid was pelleted by centrifugation at 16060 x *g* for 10 min at 4°C. The nucleic acid pellet was washed in 70% (v/v) ethanol, and left to air dry before resuspending in 10 µl nuclease free water (Qiagen). This protocol was found to be unreliable and was replaced by the protocol in section 2.2.1.3



### 2.2.1.3 Nucleic acid extraction for flow cytometry sorted PPE cells from the AMT cruise (2008)

Nucleic acid was extracted from flow cytometry sorted cells, using an adaptation of the protocol of Qiagen's DNeasy Blood & Tissue Kit (Lepère *et al.*, 2011). Sorted samples were adjusted to a vol. of 200 µl with nuclease free water. ATL buffer (180 µl) and 20 µl lysozyme (20 mg ml<sup>-1</sup>) were added and the mix was incubated for 30 min at 37°C. After incubation, AL buffer (200 µl), 25 µl proteinase K (10mg ml<sup>-1</sup>) and 20 µl glycogen were added and mixed thoroughly. Following incubation at 56°C for 30 min, absolute ethanol (200 µl) was added to the sample. The mixture was added to a DNeasy Spin column, centrifuged at 6,000 x g for 1 min and the flow through discarded. The column was washed with AW1 (500 µl) and centrifuged for 1 min at 6,000 x g, the flow through again discarded. A further wash with AW2 (500 µl) was undertaken and centrifuged for 3 min at 16,000 x g to dry the DNeasy membrane. The DNeasy spin column was placed in a clean 1.5 ml Eppendorf and AE buffer (200 µl) was added to the membrane and incubated at RT for 2 min. The nucleic acids were eluted by centrifugation for 1 min at 16,000 x g and stored at -20°C.

### 2.2.1.4 DNA extraction for total phytoplankton filters from the AMT cruise

In order to target PPEs, prefiltered seawater samples (through 10 µm and 5 µm pore size polycarbonate filters (PALL)) were further filtered onto 0.2 µm pore size Supor filters (PALL). The filters were flash frozen and stored at -80°C in cryovials containing Lysis buffer (20 mM EDTA, 400 mM NaCl, 0.75 M sucrose, 50 mM Tris-HCl pH 9.0, sterilized by autoclaving).

Filters were thawed and SDS (350  $\mu$ l, 10% w/v) and 33.5  $\mu$ l of proteinase K (recombinant from Roche, 10 mg ml<sup>-1</sup>) were added. The samples were mixed and incubated at 37°C for 30 min, followed by 10 min at 55°C. The lysate, with the filter included, was transferred into a 15ml tube (BD Biosciences) and the nucleic acids were extracted with one volume phenol:chloroform:isoamyl alcohol (25:24:1). The filter was dissolved by vortexing and the solution was centrifuged for 5 min at 6,000 x g at room temperature. The aqueous phase was removed and the nucleic acids were further extracted with one volume chloroform:isoamyl alcohol (24:1) and centrifuged for 5 min at 6,000 x g at room temperature. The aqueous phase was then placed into a fresh tube and the nucleic acid precipitated with 0.4 vol. 7.5 M ammonium acetate and one vol. isopropanol at room temperature for 10 min. The nucleic acids were precipitated by being centrifuged for 45 min at 6,000 x g. The nucleic acid pellet was washed with 70% (v/v) ethanol and the dried nucleic acid pellet was resuspended in 30  $\mu$ l of TE2 (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA pH 8.0).

## **2.2.2 Polymerase Chain Reaction (PCR)**

### 2.2.2.1 Conditions for plastid-biased 16S rRNA PCR

In order to specifically target photosynthetic picoeukaryotes a primer set specific for the plastid 16S rRNA of picoeukaryotes, designed by Fuller *et al.*, (2006a), was used. All PCR primers were obtained from Invitrogen and were purified via 'desalting'. The primer sequences for PLA491F and OXY1313R are described in Table 2.3. Reactions were undertaken in a 50  $\mu$ l volume in a T3000 Thermocycler (Biometra). The conditions for the PCR are as described overleaf.

Step 1	Step 2	Step 3
95°C – 5 min x1	95°C – 30 sec	72°C – 6 min x 1
	60°C <sup>a</sup> – 30 sec	4°C – Hold
	72°C – 45 sec <sup>b</sup>	
	} x 30	

<sup>a</sup> The temperature of this step was primer specific. <sup>b</sup> This step was carried out with the estimation of 1 min/kb of desired product.

The components of a typical 50 µl reaction are described in Table 2.4.

#### 2.2.2.2 Conditions for the amplification of picoeukaryotic nuclear 18S rDNA

In order to target the nuclear 18S rDNA of picoeukaryotes, PCR conditions and primers from Lopez-Garcia *et al.*, (2003) were used. The sequence of the two primers, 82F and 1498R, are described in Table 2.3.

Step 1	Step 2	Step 3
94°C – 2 min x1	94°C – 15 sec	72°C – 10 min x 1
	55°C <sup>a</sup> – 30 sec	4°C – Hold
	72°C – 2 min <sup>b</sup>	
	} x 30	

<sup>a</sup> The temperature of this step was primer specific. <sup>b</sup> This step was carried out with the estimation of 1 min/kb of desired product.

The components of a typical 50 µl reaction are described in Table 2.4.

**Table 2.3. PCR primer sequences used during this study. Where V = A, G or C and N = A, C, G or T.**

Primer	Target	Sequence (5' to 3')	Ref.
PLA491F	Marine-algal plastid 16S rRNA gene	GAG GAA TAA GCA TCG GCT AA	Fuller <i>et al.</i> , 2006a
OXY1313 R	Marine-algal plastid 16S rRNA gene	CTT CAC GTA GGC GAG TTG CAG C	West <i>et al.</i> , 2001
82F	Nuclear 18S rRNA gene (all eukaryotes)	GAA ACT GCG AAT GGC TC	Lopez-Garcia <i>et al.</i> , 2003
1498R	Nuclear 18S rDNA (all eukaryotes)	CAC CTA CGG AAA CCT TGT TA	Lopez-Garcia <i>et al.</i> , 2003
Pry421F	Nuclear 18S rDNA (Prymnesiophyte-specific)	AGC AGG CGC GTA AAT TGC CCG	Jardillier (unpublished)
Pry1572R	Nuclear 18S rDNA (Prymnesiophyte-specific)	TCA ACG YRC GCT GAT GAC A	Jardillier (unpublished)
Oligo(dT) Ran	n/a	TTT TTT TTT TTT TTT TTT TTT TTT VNN	This study

#### 2.2.2.3 Conditions for the specific amplification of prymnesiophyte sequences

In order to specifically target prymnesiophytes a nested PCR was used. The first round of amplification was undertaken using the nuclear 18S rDNA protocol, as described above in section 2.2.2.2. 1µl of the product, from this reaction, was transferred to a second round of amplification, using the standard conditions overleaf.

**Step 1**

94°C – 2 min x1

**Step 2**

94°C – 15 sec

63°C -1°C per cycle<sup>a</sup> – 30 sec } x 1072°C – 2 min<sup>b</sup>**Step 3**

94°C – 15 sec

60°C<sup>a</sup> – 30 sec +1 sec per cycle } x 2572°C – 2 min<sup>b</sup>**Step 4**

72°C – 7 min

4°C – Hold

<sup>a</sup> The temperature of this step was primer specific. <sup>b</sup> This step was carried out with the estimation of 1 min/kb of desired product.

The components utilised for this PCR are described in Table 2.4 and used the primers Pry421F and Pry1572R (Table 2.3) (Jardillier pers. comms.).

**Table 2.4: The volumes of components used (final concentrations) for the specific PCR reactions.**

Component	Marine-algal plastid 16S rDNA	Nuclear 18S rDNA	Prymnesiophyte-specific 18S rDNA
DNA	1µl	1µl	1µl
dH <sub>2</sub> O	29.5µl	34.5µl	34.5µl
10x Buffer	5µl	5µl	5µl
MgCl <sub>2</sub>	2.5µl (2mM)	1.5µl (1.5mM)	1.5µl (1.5mM)
Bovine Serum Albumin	2.5µl (0.05mg)	2.5µl (0.05mg)	2.5µl (0.05mg)
Forward primer	4µl (800nM)	2µl (400nM)	2µl (400nM)
Reverse primer	4µl (800nM)	2µl (400nM)	2µl (400nM)
dNTPs	1µl (200 µM each)	1µl (200 µM each)	1µl (200 µM each)
Recombinant Taq (Invitrogen)	0.5µl (2.5U)	0.5µl (2.5U)	0.5µl (2.5U)

### 2.2.3 PCR purification

Products from PCR reactions were run on 1% (w/v) agarose gels, stained with EtBr (10mg/mL stock) and viewed under a UV transilluminator (Flowgene), prior to the correct size band being excised from the gel. Purification was achieved using Qiagen's Gel Purification kit. Briefly the gel slice was melted in the manufacturer's buffer at 55°C. After the gel had been melted, isopropanol was added, before the solution was placed on a column and centrifuged. The column was then washed with the appropriate buffer and centrifuged. Remaining ethanol was removed from the column and the DNA was eluted in 30-50 µl water.

### 2.2.4 Construction of clone libraries

#### 2.2.4.1 Construction of clone libraries using DNA extracted from sorted PPE cells from SOLAS cruise D326

Purified PCR product was added to 1 µl pCR2.1<sup>®</sup>-TOPO<sup>®</sup> vector (Invitrogen), plus 1 µl of the supplied salt solution and incubated at room temperature for 5 min. The ligation reaction (2 µl) was then added to 50 µl *E.coli* TOP10F' chemically competent cells and incubated on ice for 30 min. Cells were heat shocked (42°C) for 30 sec and immediately transferred back to ice (2 min). Room temperature SOC medium (section 2.1.1) (250 µl) was added and incubated under agitation (240 rpm) in an orbital shaker (New Brunswick Scientific) for 90 min at 37°C. 1.5% (w/v) LB agar plates, with the addition of ampicillin (100 µg ml<sup>-1</sup>), were used for plating. The plates contained 100 µl IPTG (100 mM) and 20 µl XGal (50 mg ml<sup>-1</sup>) spread on them, when required for blue/white screening. The plates were incubated for 30 min at 37°C, prior to dilutions of the sample being spread onto them and incubated at 37°C overnight. White colonies were individually picked and re-plated on

LB/Amp/IPTG/XGal plates and incubated overnight. Subsequently, positive (white) colonies were resuspended in 10  $\mu$ l nuclease free water. 1  $\mu$ l was used for PCR reactions, using the plastid biased 16S rRNA primers (PLA491F and OXY1313R).

#### 2.2.4.2 Construction of clone libraries from DNA extracted from AMT18 samples

The construction of clone libraries from the DNA extracted from AMT18 samples was undertaken using the pGEM® T Easy Vector system (Promega). Approximately 100 ng of purified PCR product was added to 1  $\mu$ l of pGEM T Easy vector (50 ng  $\mu$ l<sup>-1</sup>), 1  $\mu$ l of T4 DNA ligase (3U  $\mu$ l<sup>-1</sup>) and 5  $\mu$ l of 2X Rapid ligation buffer. In order to obtain the maximum number of transformants the reaction was incubated at 4°C overnight. An addition of 2  $\mu$ l of the ligation reaction to 25  $\mu$ l *E. coli* JM109 High Efficiency Competent cells (Promega) was undertaken and incubated on ice for 20 mins. Cells were heat shocked for 50 sec at 42°C and immediately placed back in ice for a further 2 min. Room temperature SOC medium (see section 2.1.1) (975  $\mu$ l) was added and incubated under agitation (240 rpm), in an orbital shaker (New Brunswick Scientific), for 90 min at 37°C. LB plates, with the addition of ampicillin (100  $\mu$ g ml<sup>-1</sup>), were used for plating. The plates contained 100  $\mu$ l IPTG (100 mM) and 20  $\mu$ l XGal (50 mg ml<sup>-1</sup>) spread on them, when required, for blue/white screening. The plates were incubated for 30 min at 37°C, prior to dilutions of the sample being spread onto them and incubated at 37°C overnight. White colonies were individually picked, resuspended in 200  $\mu$ l LB together with ampicillin (100  $\mu$ g ml<sup>-1</sup>) and incubated overnight. Subsequently 1  $\mu$ l of cell suspension was used for PCR reactions.

### 2.2.5 Restriction endonuclease digestion

Restriction enzyme digestion of DNA fragments was undertaken according to the manufacturer's protocols (Fermentas). Briefly, 10U of the restriction enzyme was used in a reaction volume of 30  $\mu$ l with the recommended buffer. For double digests the restriction enzymes were mixed together and the digest made simultaneously. The reactions were incubated for at least 1 h at 37°C, prior to being analysed by agarose gel electrophoresis on a 2% (w/v) gel (stained with EtBr) and visualised on a UV transilluminator.

### 2.2.6 Sequencing of clone library inserts

PCR products of clones with unique restriction fragment length polymorphism (RFLP) patterns, were sent to the NERC Biomolecular Analysis Facility in Edinburgh for Sanger sequencing, using an ABI 3730 instrument.

## 2.3 Transcriptomic techniques

### 2.3.1 Sample collection and culture maintenance

#### 2.3.1.1 *Ochromonas* sp. (CCMP 584) culture conditions

A strain of *Ochromonas* (CCMP 584), originally isolated from the Sargasso Sea, was cultivated in 0.5 L f/2 medium (Table 2.2). A grain of rice was added, as a food source for this mixotrophic alga, to enhance the growth of heterotrophic bacteria (Andersen, 2005). The culture was grown in a 12h light/dark cycle at a constant temperature of 21°C. The light intensity was set at 30  $\mu$ E m<sup>-2</sup> s<sup>-1</sup>. Cultures were maintained in a volume of 0.5 L, in a plastic flask, with a 0.2  $\mu$ m filter lid (BD Biosciences) for increased air transfer.



### 2.3.1.2 Clade VIIA prasinophyte (RCC1124) culture conditions

RCC1124, a clade VIIa prasinophyte, originally isolated by L. Jardillier from the North Atlantic (48° 50'N, 16° 29'W), was maintained in K Medium (Table 2.2), at a constant temperature of 21°C under a 12h light/dark cycle, with a light intensity of 30  $\mu\text{E m}^{-2} \text{s}^{-1}$ . Cultures were maintained in a volume of 0.5 L, in a plastic flask, with a 0.2  $\mu\text{m}$  filter lid (BD Biosciences) for increased air transfer.

### 2.3.1.3 Collection of environmental RNA samples

In order to remove larger particulates, 20 L water from the CTD was pre-filtered with a 100  $\mu\text{m}$  mesh (Millipore). Pluronic was added to a final concentration of 0.05% (w/v) to avoid adhesion of cells to the carboys and tubes. In order to target picoeukaryote cells, these were size fractionated through 10 and 5  $\mu\text{m}$  filters (Millipore), before being collected in a Cell Trap™ (MEM-TEQ, Wigan, UK) with a 0.2 $\mu\text{m}$  filter. Approximately 4 L of water was washed through each Cell Trap™ with triplicates taken at each station. Cells were washed off the filter, according to the manufacturer's instructions, before being placed in a cryovial and immediately flash frozen using liquid nitrogen and stored at -80°C.

### 2.3.1.4 Flow cytometry sorting of environmental transcriptomic samples

Flow cytometry was used to target PPEs. Work was performed at the National Oceanographic Centre, Southampton, with the help of Ross Holland. PPEs were targeted using their endogenous chlorophyll fluorescence (FL3) and side scatter (SSC) as distinguishing factors, on a MoFlo flow cytometer (Dako-Cytomation). To limit the degradation of the RNA, small volumes were defrosted systematically and

sorted directly into acid phenol. Depending on the sample, between 250,000 and 1 million cells were sorted and immediately flash frozen and stored at -80°C.

### 2.3.2 Total RNA extraction

Subsamples of cultures (50ml) were harvested at the beginning of the light cycle. The *Ochromonas* cultures were pre-filtered on a 100 µm mesh (Millipore) to remove large debris. Pluronic was added, at a final concentration of 1 % (w/v) and the cells centrifuged at 6,000 x *g* for 15 min.

Total RNA was extracted using a Ribolysing technique, adapted from Gilbert *et al.* (2000). Following centrifugation cells were resuspended in 200 µl of solution I (0.3M sucrose, 0.01M sodium acetate) and 200 µl of solution II (2% (w/v) sodium dodecyl sulphate (SDS), 0.01M sodium acetate). An equal volume of acid phenol was added to the suspension and the solution added to Lysing Matrix E tubes (MP Bio). Cells were lysed in a Hybaid Ribolyser at a speed setting of 6.5 for 45 sec, before being immediately placed on ice. The lysed cells were centrifuged for 5 min at 16,000 x *g* and the aqueous phase transferred to a fresh tube. Acid phenol (400 µl) was added to the aqueous phase and incubated at 65°C for 4 min, followed by 10 sec on dry ice ethanol. There followed 5 min of centrifugation at 16,000 x *g*, the aqueous phase was transferred to a fresh tube and 400 µL phenol chloroform added. After a further round of centrifugation the aqueous phase was transferred to a fresh tube. RNA was precipitated using 2.5 volumes absolute ethanol, 1/10 volume sodium acetate and placed at -20°C overnight. The RNA was pelleted by centrifugation (16,000 x *g*) for 30 min at 4°C, washed with 70% (v/v) ethanol, air dried and resuspended in 10 µl of nuclease free water (Qiagen).

### 2.3.3 mRNA purification for cultures

Eukaryotic mRNA was purified from total RNA and bacterial mRNA using poly (dT) magnetic beads (NEB). These beads preferentially bind covalently to the poly (A) tail present on eukaryotic mRNA. Total RNA was added to 500 µl of Lysing/Binding buffer (100 mM Tris-HCl, pH 7.5, 500 mM LiCl, 0.5% LiDS, 1 mM EDTA, 5mM DTT) and incubated for 5 min at room temperature with gentle agitation. This solution was added to the magnetic beads, as per the manufacturer's instructions and incubated with gentle agitation for 10 min. Using a Neodymium magnet the beads were attracted to the side of the tube and the solution removed. The beads were washed according to the manufacturer's protocols using the appropriate buffers, before the mRNA was eluted using 100 µl of elution buffer (20 mM Tris-HCl, pH 7.5, 1 mM EDTA) (NEB). A second round of purification was undertaken to minimise the amount of ribosomal RNA in the sample.

### 2.3.4 Enrichment of mRNA from sorted samples

Eukaryotic mRNA was enriched using Terminator<sup>TM</sup> 5'-phosphate dependent exonuclease (Epicentre). This is a 5'→3' processive exonuclease that digests RNA with a 5' monophosphate. Eukaryotic mRNA, which possesses a 5' cap, is protected from digestion and is thus enriched in the sample. 1U of the terminator exonuclease was added, with the appropriate buffer, to the total RNA, along with an RNase Inhibitor (Promega). The mixture was incubated at 30°C for 60 min in a thermocycler (Biometra), with a heated lid to stop evaporation. The reaction was terminated by a phenol extraction, using buffer saturated phenol (pH 4.5). RNA was precipitated using 0.1 volumes of sodium acetate, plus 2.5 volumes of absolute ethanol and incubated on dry ice for 30 min. RNA was pelleted by centrifugation at

4°C (16,000 x *g*) for 30 min, washed once with 70% (v/v) ethanol and resuspended in RNase free water.

### 2.3.5 cDNA library construction for mRNA from cultures

Purified mRNA was incorporated into a cDNA library, using Invitrogen's Cloneminer II cDNA library construction kit. The purified mRNA was used as a template for the synthesis of first strand cDNA. A biotin-attB2-Oligo (dT) primer targeted the poly (A) tail of the eukaryotic mRNA and Superscript<sup>TM</sup> III reverse transcriptase was used to create the first strand, as per the manufacturer's protocols. The second strand of cDNA was synthesised using *E. coli* DNA ligase, *E. coli* DNA polymerase I and *E. coli* RNase, using the buffers and incubation times as detailed in the manufacturer's protocol. The cDNA was blunt ended using T4 DNA polymerase, to allow the ligation of an attB1 adapter to the 5' end. The attB1 adapter was prevented from ligating to the 3' end by the presence of Biotin. cDNA was size fractionated, using column chromatography to remove primers and short cDNAs. During the BP reaction, Gateway Technology<sup>TM</sup> was used to facilitate the recombination of the attB flanked cDNA into a plasmid containing attP sites (pDONR 222). The plasmid was transformed into ElectroMax DH10B T1 phage resistant cells, using a BioRad Gene pulser II electroporator, as per the manufacturer's handbook. The kanamycin resistant cassette on the plasmid was expressed by growing the electroporated cells in SOC medium (see section 2.1.1) for one hour at 37°C with shaking (240 rpm). Dilutions were spread on LB plates, containing 50 µg ml<sup>-1</sup> kanamycin and incubated overnight.

Approximately 200 colonies were picked off the LB + kanamycin plates and resuspended in 10 ml of LB + kanamycin (50 µg ml<sup>-1</sup>) and incubated overnight in a

shaking incubator (240 rpm) at 37°C. The plasmid was extracted with Qiagen's Miniprep kit, using the supplied columns and buffers, according to the manufacturer's instructions. The plasmid was eluted in 50 µl of elution buffer (Qiagen). Eluted plasmid was digested, using the restriction enzyme *BsrGI* (NEB) for 2 h at 37°C. The digested plasmid was run on a 2 % (w/v) agarose gel, before inserts between 250bp and 1.5kb, were excised from the gel and extracted as described in section 2.2.3. Inserts above 2kb were not extracted from the gel, in order to avoid the plasmid backbone, which is 2.2kb in size.

### **2.3.6 mRNA amplification from sorted samples**

Amplification of the mRNA was undertaken using Ambion's MessageAmp<sup>TM</sup> II aRNA amplification kit. A T7 Oligo (dT) primer (Table 2.3) was used to specifically prime the poly(A) tail of the mRNA, acting as the basis for the synthesis of first strand cDNA. Second strand cDNA was synthesised utilising DNA polymerase and RNase H, as per the manufacturer's instructions. The cDNA was purified using a filter cartridge and eluted in nuclease free water at 55°C. aRNA was synthesised using a T7 enzyme mix at 16°C for 20 h, before being purified using an aRNA filter cartridge and eluted in nuclease free water. A second round of amplification was undertaken according to the manufacturer's protocols. First strand cDNA was synthesised from the aRNA using second round primers (see below). Subsequently, second round cDNA was synthesised using the T7 Oligo (dT) primer. The remainder of the second round of amplification was the same as the first round.

First strand cDNA was created using a Transcriptor first strand cDNA synthesis kit (Roche). A final concentration of 60 µM of random hexamer primer was added to the aRNA and incubated at 65°C for 10 min in a thermocycler (with a heated lid).

The reverse transcription mix was then added, which consisted of 10U of Transcriptor Reverse Transcriptase (Invitrogen), 1x Transcriptor Reverse Transcriptase reaction buffer (Invitrogen), 20U of Protector RNase Inhibitor (Invitrogen) and 1mM each of dNTPs (Invitrogen). The manufacturer's protocol, for obtaining cDNA up to 4kb in length, was followed. The mix was incubated at 25°C for 10 min, followed by a 30 min incubation at 55°C. The enzymes were inactivated by incubation at 85°C for 5 min.

Second strand synthesis used the enzymes DNA polymerase I, RNase and DNA ligase, along with dNTPs (1 mM each) and 10X second strand buffer to create the second strand, with an incubation time of 2 h at 16°C. The reaction was inactivated using 10 µl EDTA.

A poly(A) tail was added to the 3'-OH end of the cDNA, using terminal deoxynucleotidyl transferase (Fermentas). The reaction was undertaken in a volume of 20 µl, consisting of cDNA, 4 µl 5X reaction buffer, 130 pmol dATP and 30U terminal deoxynucleotidyl transferase (Fermentas). The mixture was incubated for 15 min at 37°C before the enzyme was inactivated by heating to 70°C for 10 min.

The poly(A) tailed cDNA was amplified using an Oligo (dT) primer: three degenerate (either A, C or G) bases were added to the 3' end of the primer to add specificity to the primer for the cDNA end of the poly(A) tail. The PCR was undertaken using recombinant Pfu DNA polymerase (Fermentas). This consisted of 5 µl of 10X buffer with MgSO<sub>4</sub> (2.5 mM), 200 µM dNTPs each (final conc.) and 2.5U of Pfu DNA polymerase (a final concentration of 800nM). PCR conditions were an initial denaturation at 95°C for 2 min, followed by 30 cycles of 95°C for 30

sec, 50°C for 30 sec and 72°C for 4 min. A final extension period of 10 min at 72°C took place after the completion of the cycles.

### 2.3.7 Sequencing of cDNA using 454 technology

Size fractionated inserts were pooled together and ~5 µg of cDNA was sent to the NERC Biomolecular Analysis Facility in Liverpool. The inserts were sequenced on a Next Generation Roche 454 FLX sequencer. Data was outputted as raw reads with a corresponding quality file for the sequencing. Reads were also assembled into contigs using the Newbler assembler software (Roche).

## 2.4 Statistical analysis

### 2.4.1 Species richness estimations

#### 2.4.1.1 Non parametric estimators of species richness

The Chao1 estimate is based on the number of rare species in the sample (Colwell and Coddington, 1994).

$$S_{Chao1} = S_{obs} + \left( \frac{F_1^2}{2F_2} \right)$$

Where  $S_{obs}$  = the number of observed species in the sample;  $F_1$  = the number of observed species represented by a single individual (singleton);  $F_2$  = the number of observed species represented by two individuals (doublets).

The measure is a function of the ratio between singletons and doublets. As this ratio increases the estimate will exceed the observed species by greater margins. When all species are represented by at least two individuals then the inventory can be considered complete.

#### 2.4.1.2 Coverage values of clone libraries

An estimation of the percentage of OTUs recovered in the clone library was obtained using the equation (Good, 1953):

$$C = 1 - \frac{n_i}{N}$$

Where  $n_i$  is the number of singletons observed in the sample and  $N$  is the number of clones examined in the sample.

#### 2.4.1.3 Species diversity estimation using Margalef's diversity index

Margalef's diversity index ( $D_{Mg}$ ) takes into account the sampling effort when assessing the species richness of the assemblage measured (as described in Magurran, 2003).

$$D_{Mg} = \frac{(S - 1)}{\ln N}$$

Where  $S$  = the number of species recorded;  $N$  = the total number of individuals in the sample.

#### 2.4.1.4 Comparison of species composition with Morisita-Horn

Comparisons between the species composition of different stations was undertaken using the Morisita-Horn equation (described in Magurran, 2003).

$$C_{MH} = \frac{2\sum(a_i \cdot b_i)}{(d_a + d_b) * (N_a * N_b)}$$



Where  $N_a$  = the total number of individuals at site A;  $N_b$  = the total number of individuals at site B;  $a_i$  = the number of individuals in the  $i$ th species in A;  $b_i$  = the number of individuals in the  $i$ th species in B; and  $d_a$  (and  $d_b$ ) are calculated as follows:

$$d_a = \frac{\sum a_i^2}{N_a^2}$$

#### 2.4.1.5 Hierarchical cluster analysis

Hierarchical cluster analysis was undertaken using the R statistical program (<http://www.r-project.org/>). Dendrograms were created by a recursive process in which the pairwise similarity between every sample and every other sample is calculated. Distance calculations between samples were calculated using the Euclidean distance equation below:

$$d_{A,B} = \sqrt{\sum_{n=1}^N (A_n - B_n)^2}$$

The samples which are most similar are linked by a branch. The process is then repeated, with the samples already linked agglomerated in such a way as to be considered as a single sample. The process is repeated until all samples are joined together.

#### 2.4.1.6 Canonical correspondence analysis (CCA) and Variation Partitioning Analysis (VPA)

To explain the distribution of PPEs, canonical correspondence analysis (CCA) (Ter Braak, 1986) was undertaken utilising the Vegan package in the software program R

(<http://www.r-project.org/>). Forward CCA analysis was used to identify the factors which were significant in explaining the variation. These factors were then used in Variation Partitioning Analysis (VPA) (Borcard *et al.*, 1992). The variables were partitioned as detailed where appropriate. VPA was undertaken utilising the Vegan package in the software program R.

## **2.4.2 Phylogenetic analysis**

### 2.4.2.1 Sequence assembly

Forward and reverse rRNA reads were assembled using Seqman (DNASTAR) and alignments were manually checked to remove any ambiguous bases.

### 2.4.2.2 Chimera check

The plastid biased 16S rRNA gene sequences were checked for chimeras using the Bellerophon program (Huber *et al.*, 2004) against the greengenes 16S rRNA database ([http://greengenes.lbl.gov/cgi-bin/nph-bel3\\_interface.cgi](http://greengenes.lbl.gov/cgi-bin/nph-bel3_interface.cgi)). Resulting chimeras were subsequently removed from further analysis. The nuclear 18S rDNA sequences were checked for chimeras using KEYDNATOOLS (Guillou *et al.*, 2008), a tool to specifically detect chimeras in the eukaryotic SSU.

### 2.4.2.3 Phylogenetic analysis

Sequences were imported into a laboratory curated ARB database (Ludwig *et al.*, 2004). ARB phylogenetic trees were constructed from sequences >1200 nucleotides in length using the Jukes-Cantor correction. Sequences were aligned and inserted into existing neighbour-joining phylogenetic trees using ARB parsimony. Bootstrap analysis was undertaken with the MEGA5 program (Tamura *et al.*, 2011) using both

neighbour joining and maximum likelihood algorithms with pairwise deletion and 100,000 replications.

### **2.4.3 Transcriptomic analysis**

#### 2.4.3.1 Construction of contigs from raw 454 reads

Sequencing raw reads were compiled into contigs using the Newbler assembly program (Roche). In the case of the *Ochromonas* and environmental samples and Meta-assembler, a combination of 7 commercially available assembly programs developed for the Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis (CAMERA) project (<https://portal.camera.calit2.net>) was also used.

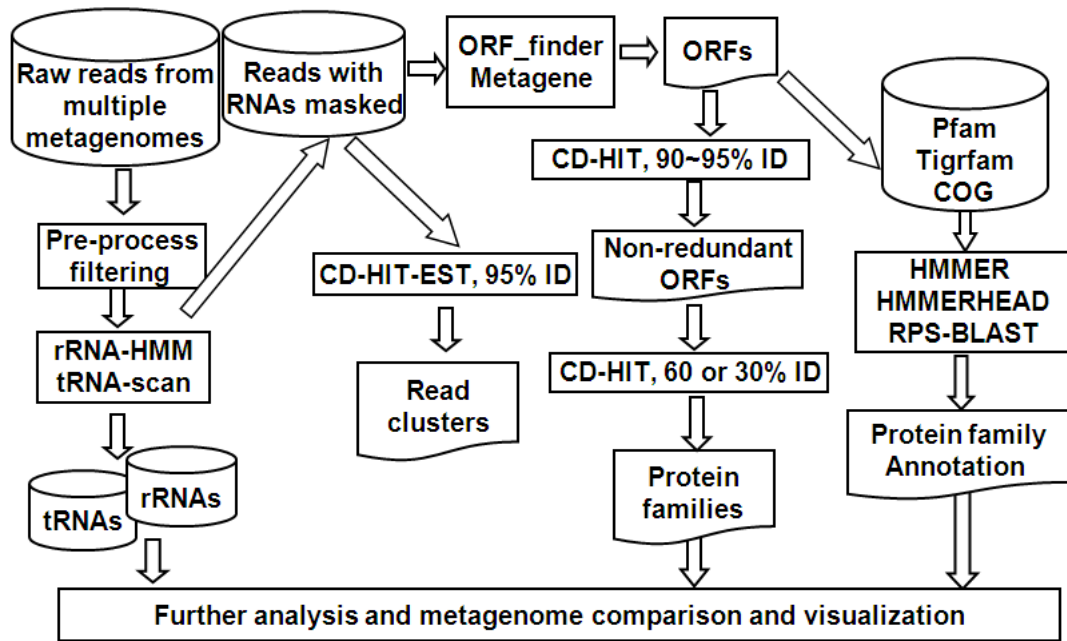
#### 2.4.3.2 Analysis of transcriptomic reads against the NCBI nr database

All contigs were annotated, using BLASTx against the NCBI nr database (Altschul *et al.*, 1990).

Phylogenetic analysis of identified rRNA reads was undertaken using BLASTn against the NCBI nr database.

#### 2.4.3.3 Rapid Analysis of Multiple Metagenomes with a Clustering and Annotation Pipeline (RAMMCAP)

The RAMMCAP pipeline (Figure 2.1) was developed by Li (2009) in order to enable the rapid analysis of large metagenomic datasets. This pipeline was developed for the CAMERA project (Seshardi *et al.*, 2007). The contigs created by the assembly programs (see section 2.4.3.1) were used as the input for RAMMCAP.



**Figure 2.1: RAMMCAP pipeline for the analysis of metagenomic datasets.**

rRNA reads were identified using a hidden Markov model (HMM) developed to have a high sensitivity and specificity for short reads (Huang *et al.*, 2009). The resulting rRNA reads were removed from further analysis.

Exact or nearly identical duplicates were identified from the input contigs using Cluster Database-High Identity with Tolerance (CD-HIT)\_EST with sequence identity of 95% over 80% of length of short reads (Li and Godzik, 2006).

Open reading frames (ORFs) were detected using the Metagene program (Noguchi *et al.*, 2006). Within each reading frame an ORF starts at the beginning of a read or at the first ATG after a stop codon. ORFs end at the first stop codon or at the end of the read. Due to the short read lengths a predicted ORF may be a portion of a complete ORF (Noguchi *et al.*, 2006). ORFs were clustered using CD-HIT firstly at 90% to identify non redundant sequences which are then further clustered at a conservative threshold with 60% identity over 80% of the length of the ORF.

ORFs were annotated against the protein families (Pfam) database (Fin *et al.*, 2008) and Tigrfam database (Haft *et al.*, 2003) with Hidden Markov Models (HMMER) (Eddy, 1998). Hits had an e-value cutoff value of 0.001.

#### 2.4.3.4 Annotation of ORFs using Cluster of orthologous groups for eukaryotic genomes (KOG)

Functional prediction of ORFs against the KOG database (Tatusov *et al.*, 2003) was undertaken using rpsblast (Altschul *et al.*, 1990).

#### 2.4.3.5 Annotation of ORFs using Kyoto Encyclopedia of Genes and Genomes (KEGG)

Functional assignment of ORFs to the KEGG database (Kanehisa and Goto, 2000) was undertaken using BLAST. The assigned genes are linked into a set of metabolic pathways.

#### 2.4.3.6 Metagenome Analyze (MEGAN)

BLASTx outputs (obtained as described in method section 2.4.3.2) were analysed using MEGAN. Taxonomic affiliations of BLASTx results were obtained using a lowest common ancestor algorithm (Huson *et al.*, 2007).

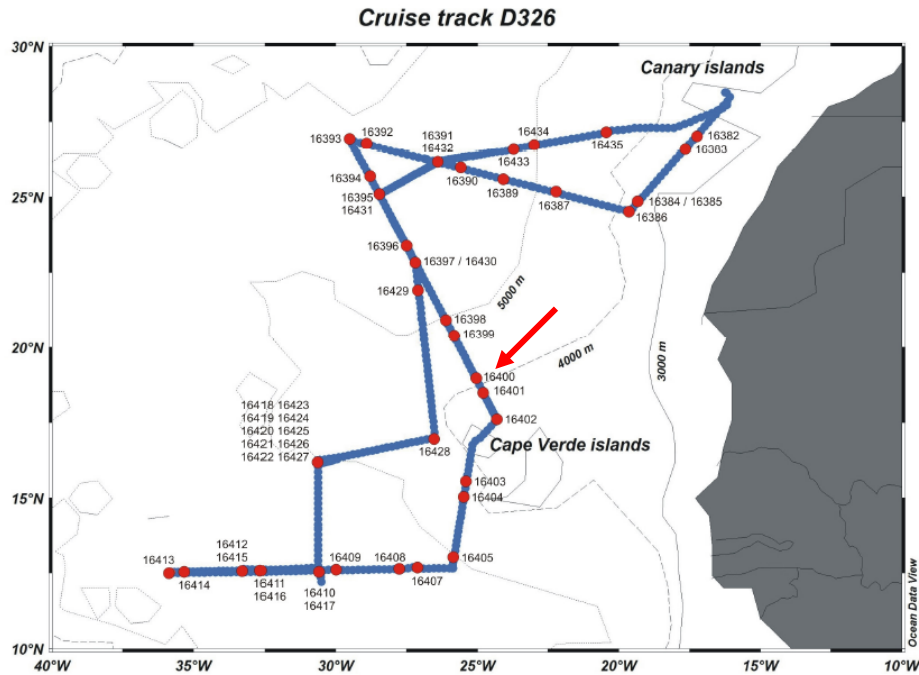
## **Chapter 3**

### **Significant CO<sub>2</sub> fixation by small prymnesiophytes in the Northeast Atlantic Ocean – the SOLAS cruise transect**

### **3.1. Introduction**

The Northeast Atlantic Ocean forms part of the Atlantic coastal and Atlantic trade wind biome, using the nomenclature of Longhurst (2007). The subtropical section of the Northeast Atlantic is regularly exposed to high Saharan dust inputs, averaging over 5g/m<sup>2</sup>/year (Jickells *et al.*, 2005). The input of dust means that there is an increased amount of bioavailable iron and other nutrients in the system, possibly impacting on metabolic processes involved in primary production, the rate of conversion, by photosynthesis, of inorganic forms of carbon to organic forms. To enhance our knowledge of atmospheric transport, deposition and cycling of dust imports on the microbial community in surface waters, the SOLAS (Surface Ocean – Lower Atmosphere Study) program was started. In order to achieve this aim a cruise (D326) was undertaken in the subtropical Northeast Atlantic in January and February 2008 (Figure 3.1) encompassing the North Atlantic Tropical Gyre (NATR) and Canary coastal upwelling (CNRY) regions (Longhurst, 2007).

Primary production is often measured using radiolabelled NaH<sup>14</sup>CO<sub>3</sub> (Li, 1994). The contribution of different communities of phytoplankton to primary production has been extensively studied before, using size fractionation approaches. The problem with size fractionation approaches is that taxonomic differentiation cannot be undertaken. In the size fraction <2-5µm, which has been shown to significantly contribute to primary production, both prokaryotes and eukaryotes can be observed (Li, 1994).



**Figure 3.1: Cruise track of D326 in the Northeast Atlantic showing the cruise track and station codes. The red arrow shows the station which was sampled for the clone library.**

Recently, radiolabelled  $^{14}\text{C}$  techniques have been combined with flow cytometry, in order to differentiate which groups of organisms contribute significantly to CO<sub>2</sub> fixation (Li, 1994; Jardillier *et al.*, 2010). Flow cytometric analysis enables the sorting of two prokaryotic genera (*Prochlorococcus* and *Synechococcus*) and a diverse grouping of unicellular algae. In a recent study in the subtropical Northeast Atlantic (Jardillier *et al.*, 2010), *Prochlorococcus* was numerically the most abundant picophytoplankton, whilst the eukaryotic fraction was about 2 orders of magnitude less abundant. Despite their low abundance eukaryotes still contributed significantly to CO<sub>2</sub> fixation (up to 40%), due to much higher cell-specific rates of carbon fixation, up to two orders of magnitude higher in the eukaryotes compared to the prokaryotic genera (Jardillier *et al.*, 2010). Recent experiments have started to



reveal the diversity of the photosynthetic picoeukaryotic fraction as a whole, showing a domination of Prymnesiophyceae, Prasinophyceae and Chrysophyceae in marine environments (Fuller *et al.*, 2006a, Not *et al.*, 2008). The eukaryotic fraction can be further discriminated on a flow cytometer into two dominant groups (Figure 3.2). The Euk-A fraction consists of cells <2µm while the slightly larger Euk-B fraction comprises cells 2-3µm in diameter. Jardillier *et al.*, (2010) showed that despite the very low abundance of cells in the Euk-B size fraction, they contributed on average 25% of the carbon fixation during the SOLAS cruise compared to 13% contribution by Euk-A. However, the taxonomic diversity of these small eukaryotic cells is still poorly understood and there has been no study directly linking the taxonomic affiliation of these cells and <sup>14</sup>C primary production measurements.

Despite the eukaryotic fraction of the picophytoplankton being shown to be a major contributor to carbon fixation, especially in the Euk-B size fraction, relatively little is known about the diversity of the eukaryotes in this size fraction. The aim of this study was to investigate the diversity of both size fractions of the eukaryotic picophytoplankton, in order to gain a greater understanding of the major players in marine primary production.

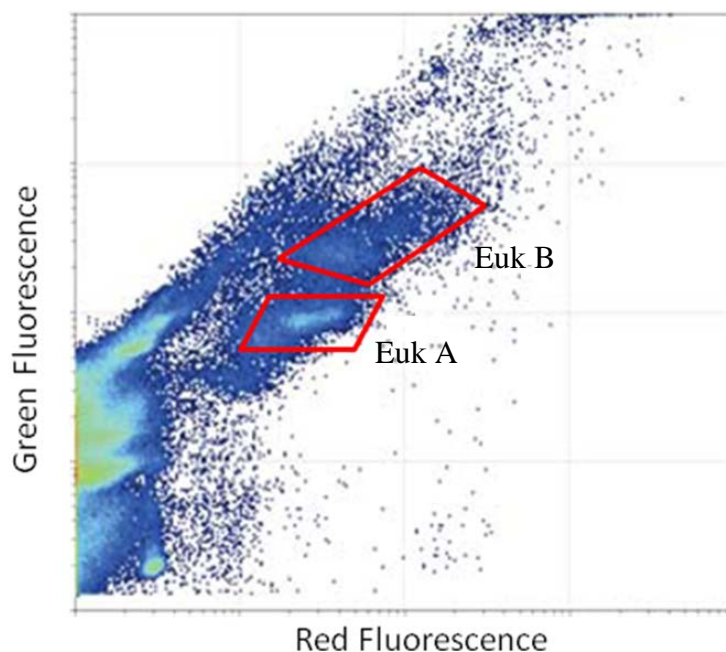
## **3.2. Materials and methods**

### **3.2.1. Collection of samples from the D326 cruise**

The specific seawater sample analysed in this work was collected at station 16400 during the D326 cruise, which took place between the 12<sup>th</sup> January and 1<sup>st</sup> February 2008 (Figure 3.1). Seawater samples from this cruise were collected at dawn in 20L Niskin bottles mounted on a conductivity-temperature-fluorescence-depth profiler

with a metallic frame. Seawater samples (5 L) were amended with 0.05% (w/v) pluronic (Sigma) and filtered through 0.2 µm pore sized CellTrap™ units (MEM-TEQ Ventures Limited, Wigan, UK). Cells were washed off, according to the manufacturer's instructions and fixed for 1 h with 1% (w/v) paraformaldehyde (PFA), flash frozen with liquid nitrogen and stored at -80 °C.

Prior to flow sorting, cells were slowly defrosted on ice and stained with SYBR Green I (Marie *et al.*, 1997). Picoeukaryote cells were discriminated using their endogenous chlorophyll *a* signal as well as stained DNA fluorescence and side scatter. Using this method, two fractions of picoeukaryotes could be distinguished (Figure 3.2) and 30,000 cells of each group were flow cytometrically sorted (FacSort flow cytometer (Becton Dickinson, US) in single sort mode at low speed (1,000 particles s<sup>-1</sup> to ensure maximum purity) onto a 0.2 µm pore-sized polycarbonate filter (Whatman) and stored at -80 °C.



**Figure 3.2.** Flow cytometry scattergram depicting the PPE size fractions sorted during this study (adapted from Jardillier *et al.*, 2010).

### **3.2.2. Extraction of environmental DNA**

DNA was extracted following the procedure described in materials and methods (see section 2.2.1.2).

### **3.2.3. PCR conditions & construction of clone libraries**

The PCR conditions and the protocol for library construction are described in the materials and methods (see sections 2.2.2.1 and 2.2.4.1, respectively).

### **3.2.4 Simpsons diversity index**

$$D = \sum \frac{n_i(n_i - 1)}{N(N - 1)}$$

Where  $n_i$  = number of individuals in the  $i^{\text{th}}$  species and  $N$  = total number of individuals.

## **3.3. Results**

### **3.3.1. Phylogenetic composition of Euk-A PPEs at station 16400 of the D326 cruise.**

Of the 96 clones which were selected for the Euk-A clone library, 37 gave positive PCR results when amplified using the 16S rRNA marine plastid-biased primers of Fuller *et al.*, (2006a). Of these 37 clones, 13 were attributed to photosynthetic eukaryotes. The remaining 24 clones were accounted for by bacterial sequences or chimeric sequences. The 13 eukaryotic clones were grouped into 2 OTUs defined on sequence data.

Prymnesiophyceae were the dominant component in the library accounting for 54% of the Euk-A PPEs. All the prymnesiophyte clones clustered in Chrysochromulina B2 (OTU 1) (see Fig. 3.3).

The remaining 46% of the library comprised Prasinophyceae which were attributed to OTU 3. All the prasinophyte clones were clustered within the environmental clade VIII (see Fig. 3.4).

**Table 3.1: Number of clones relating to each OTU and the resulting contribution to the clone library.**

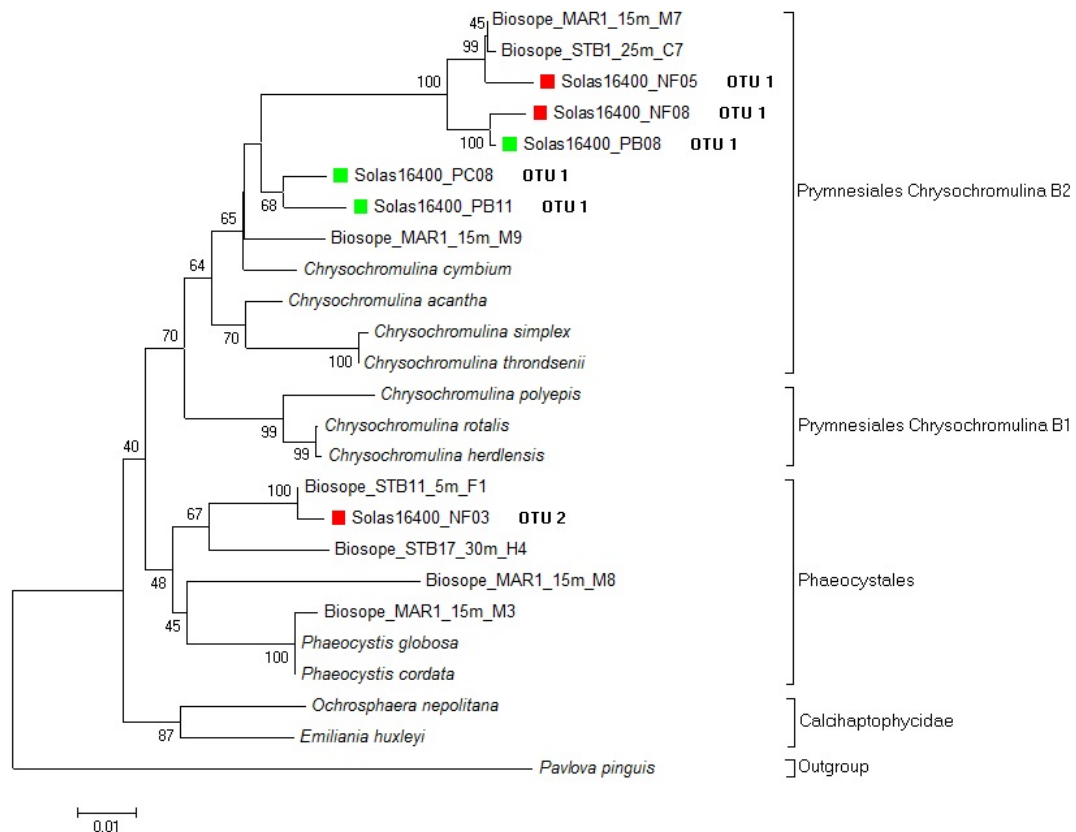
OTU	No. of Euk-A clones	Percentage contribution	No. of Euk-A clones	Percentage contribution
1	7	53.8	17	53.1
2	0	0.0	2	6.3
3	6	46.2	11	34.4
4	0	0.0	2	6.3

Estimates for alpha diversity were calculated using both Chao1 and Abundance-based coverage estimator (ACE) (see materials and methods section 2.4.1.1), which give predictions for the number of OTUs expected in the sample. Both estimators for alpha diversity indicated that the number of OTUs detected in the clone library gave a valid representation of the total diversity present in the sample (Table 3.2). However, due to the low number of clones sampled the statistical power of the tests is limited.

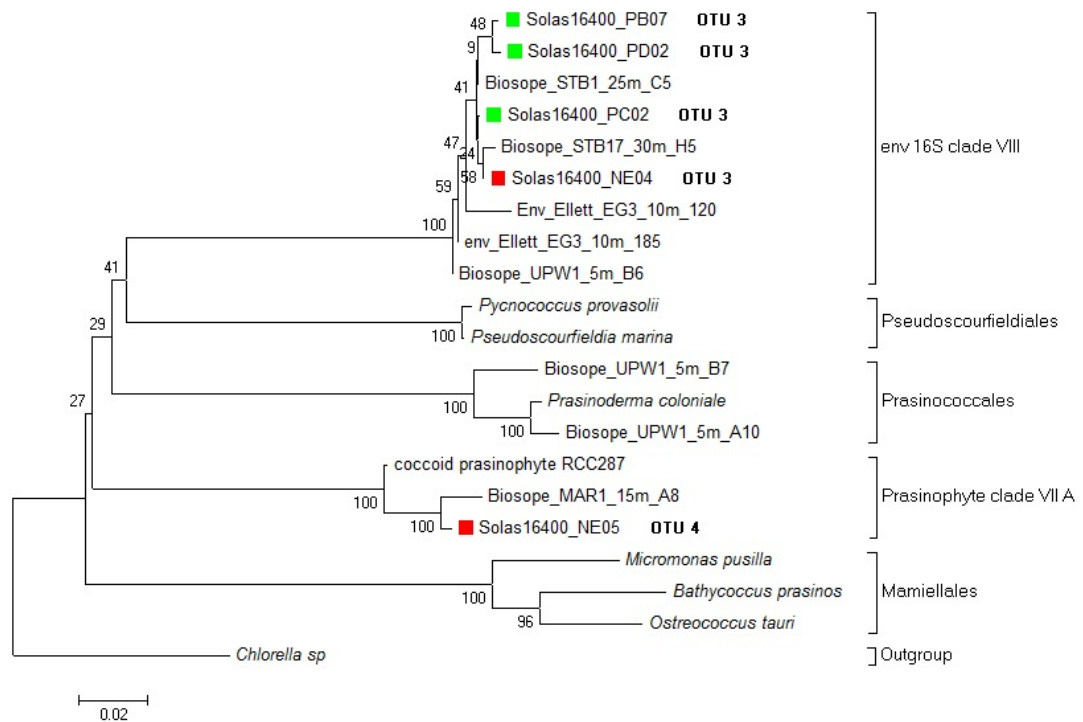
### **3.3.2. Phylogenetic composition of Euk-B PPEs at station 16400 of the D326 cruise.**

The clone library constructed from flow-sorted cells of the Euk-B fraction from cruise D326, station 16400, 5 m gave 63 positive PCR products from a total of 96

clones picked. Thirty-two of these corresponded to PPEs. Sequences related to *Acinetobacter* accounted for 22% of the clones, while other bacteria and *Synechococcus* accounted for 16% of the clones. Chimeric sequences made up the remainder of the library. In the Euk-B size fraction, 4 unique OTUs were obtained.



**Figure 3.3. Phylogenetic relationships amongst SOLAS 16S rRNA gene sequences related to the Prymnesiophyceae, obtained from the SOLAS cruise D326. Trees were created using the neighbour-joining algorithm with bootstrap values calculated from 100,000 replications. Euk-A sequences are highlighted in green and Euk-B sequences in red. The respective OTU number (see Table 3.1) follows the sequence name.**



**Figure 3.4. Phylogenetic relationships amongst SOLAS 16S rRNA gene sequences related to the Prasinophyceae, obtained from the SOLAS cruise D326. Trees were created using the neighbour-joining algorithm with bootstrap values calculated from 100,000 replications. Euk-A sequences are highlighted in green and Euk-B sequences in red. The respective OTU number (see Table 3.1) proceeds the sequence name.**

16S rRNA gene sequences related to the Prymnesiophyceae were the dominant component of the Euk-B library, accounting for 59% of the sequences. OTU 1 was the dominant prymnesiophyte clone accounting for 53% of the library and was related to *Chrysochromulina B2*. The other prymnesiophyte clones (OTU 2) clustered in the Phaeocystales clade (see Fig. 3.3 and Table 3.1). Prasinophyceae also contributed significantly to the clone library, accounting for ~34% of the PPE clones. The majority of prasinophyte sequences clustered within the prasinophyte clade VIII (OTU 3), comprising as yet only environmental sequences, whilst OTU 4

clustered within prasinophyte clade VIIA. This OTU was closely related to a cultured strain from the Roscoff culture collection (RCC287). One OTU had the closest BLAST hit to the cryptophyte species *Hemiselmis virescens*.

**Table 3.2: Number of PPE related clones in each library and the corresponding OTUs observed. Number of OTUs predicted using the diversity indicators Chao1 and ACE for each library.**

Sample	Clones	OTUs detected	Chao1	Chao1 SD	ACE
Euk-A	13	2	2	0.05	2
Euk-B	32	4	4	0.16	4.75

The 4 OTUs detected in the Euk-B clone library were a good representation in the sample of PPE diversity according to both Chao1 and ACE estimators (Table 3.2).

### 3.3.3 Diversity statistics

Margalef's diversity index, an estimation of alpha diversity showed that the Euk-B clone library had a higher diversity (0.87) compared to the Euk-A library (0.39). The Margalef index only takes into account the number of OTUs and sampling effort, whereas the Simpson Index (D) also takes into account the dominance of OTUs within the sample. D is influenced by two factors, the numbers of species present in the sample and the evenness in distribution of those species. The Simpson Index results indicate that the Euk-B size fraction has a slightly higher diversity ( $1-D=0.61$ ) than the Euk-A fraction ( $1-D=0.54$ ). In the Euk-A size fraction the species have an equitable distribution, whereas in the Euk-B size fraction half of the OTUs comprised 87% of the library.

Between the two libraries 4 unique OTUs were identified. Only two of these OTUs were observed in both size fractions. One of these OTUs (OTU 1) clustered in *Chrysochromulina* clade B2, while the other was a clade VIII prasinophyte (OTU 3).

### **3.4. Discussion**

Despite only contributing a small fraction of the phytoplankton abundance, about a third of the CO<sub>2</sub> fixation, during the SOLAS cruise, was attributable to the eukaryotic fraction of the pico-phytoplankton. In agreement with Li (1994) it was found that the larger Euk-B size fraction had almost double the contribution to carbon fixation, compared with the Euk-A fraction (Jardillier *et al.*, 2010). The significant contribution of the eukaryotic fraction of the picophytoplankton to primary production highlights the fact that investigating the diversity of these organisms, especially those in the Euk-B fraction, is important to the understanding of carbon fixation in the oceanic environment.

Prymnesiophytes have been shown to be a major contributor to the eukaryotic fraction of the pico-phytoplankton from studies in both the Atlantic and Pacific Oceans (Gibb *et al.*, 2000 and Liu *et al.*, 2009). In both size fractions prymnesiophytes were the dominant component, based on the presence of specific pigments (Gibb *et al.*, 2000) or following construction of clone libraries using the LSU rRNA gene (28S rRNA) (Liu *et al.*, 2009). Prymnesiophytes were also found to dominate FISH results from the same station in the Euk-B size fraction (Jardillier *et al.*, 2010). Recently, the extensive diversity of prymnesiophytes is beginning to be revealed, with most of the environmental sequence data, thus far, being related to *Chrysochromulina* clade B2 in the order Prymnesiales (Liu *et al.*, 2009). An OTU, related to this clade, was observed in the Euk-A size fraction, accounting for all of



the prymnesiophyte sequences in the library. The dominant prymnesiophyte OTU in the Euk-B size fraction related to *Chrysochromulina* B2. A sequence related to another major prymnesiophyte lineage, Phaeocystales, was observed in the Euk-B size fraction, in agreement with previous studies which have found sequences relating to this lineage, in the pico size fraction, in different regions of the world's oceans (Moon van der Staay *et al.*, 2000 and Liu *et al.*, 2009).

Prasinophytes comprised just under half of the Euk-A library (46%). Sequences attributed to prasinophytes in this size fraction, were clustered in clade 16S-VIII (OTU 3). This is an entirely environmental clade which was reported during the Biosope cruise in the South Pacific (Lepère *et al.*, 2009). Lepère *et al.*, (2009) found, when investigating the picoeukaryotes as a whole, that the distribution of this clade spanned a wide range of nutrient concentrations. In the Euk-B size fraction, as well as sequences related to clade VIII, one OTU clustered in clade VIIA (OTU 4), a clade which has recently had members isolated into culture (RCC287) (Guillou *et al.*, 2004). This clade has a more restricted range than members of clade 16S-VIII, being restricted to areas with elevated nutrient concentrations. FISH results on sorted cells from the same SOLAS station did not reveal the presence of prasinophytes in the sample (Jardillier *et al.*, 2010). FISH probes used by Jardillier *et al.*, (2010) targeted the Mamiellales, which have been identified as significant contributors to the picophytoplankton in coastal waters (Worden *et al.*, 2004), but these probes may underestimate the contribution of prasinophytes because the probes did not specifically target clade 16S-VIII prasinophytes.

Despite the low number of OTUs observed in the libraries (a total of 4), the Chao1 and ACE estimates, for alpha diversity suggested that the number of OTUs observed

in the library was equal to that which was predicted by the diversity indices. However, due to the limited number of PPE sequences which were obtained, the amount of PPE diversity could have been underestimated and further clones would have to be analysed to further investigate the phylogenetic composition at this station.

Of the sequences obtained in the clone libraries many had no close relatives to cultured representatives. This suggests that a large proportion of oceanographic CO<sub>2</sub> fixation is being undertaken by organisms which are poorly characterised. This highlights the need to obtain cultures of these novel organisms, so that physiological studies can be undertaken to obtain a more detailed understanding of carbon fixation in the marine environment.

This preliminary study, at a single station in the northeast Atlantic, using marine plastid-biased 16S rRNA gene primers, began to highlight the extent of PPE diversity, especially amongst the pico-prymnesiophytes. The challenge now was to move from a small scale study, as was undertaken here, to a much larger basin-scale study, so that the phylogenetic composition of the PPE community could be investigated in the two picoeukaryote size fractions across a range of trophic levels. A larger basin scale study would allow the incorporation of various environmental metadata into the analysis to attempt to understand what factors are driving the biogeography of specific PPE classes.

## **Chapter 4**

### **Community structure of flow-sorted PPE populations along an Atlantic Meridional Transect**

## **4.1 Introduction**

About half of the net global primary production on Earth is accounted for by carbon fixation in the ocean (Field *et al.*, 1998). Primary production is dominated in the oceanic environment by planktonic organisms with only a minor contribution being made by the macroalgae in coastal regions. Although both prokaryotes and eukaryotes contribute to this primary production it is becoming increasingly apparent that it is the eukaryotic component, particularly organisms less than 3µm in size, that are important in this respect.

Electron microscope counts and HPLC pigment distributions have shown that prymnesiophytes and pelagophytes are the dominant classes of the photosynthetic component of the eukaryotic picoplankton (Andersen *et al.*, 1996). Recently molecular approaches, especially the cloning and sequencing of the 18S rRNA gene from natural samples, has unveiled a considerable diversity of picoeukaryotes, with the identification of novel groups of sequences unrelated to those from cultured organisms (Moon van der Staay *et al.*, 2001; Diez *et al.*, 2001). 18S rRNA gene clone libraries constructed from a wide range of locations (see Shi *et al.*, 2009; Diez *et al.*, 2001; Massana *et al.*, 2004) have often shown a similarity in sequence affiliations (Massana and Pedros Alio, 2008; Worden and Not, 2008) with the clone libraries being dominated by non-photosynthetic lineages (Not *et al.*, 2008). When considering only sequences related to photosynthetic lineages they primarily belonged to the class Prasinophyceae (Vaulot *et al.*, 2002). However, some of these libraries revealed the presence of novel phototrophic lineages such as picobiliphytes (Not *et al.*, 2007; Hamilton *et al.*, 2008) and rappemonads (Kim *et al.*, 2010). In contrast to 18S rRNA gene clone libraries, epifluorescence microscopy typically

revealed a dominance of photosynthetic cells over heterotrophic cells in marine environments (Masquelier and Vaulot, 2008). This suggests that 18S rRNA gene clone library studies may give a biased view of diversity especially regarding photoautotrophs and has led to alternative approaches to targeting photosynthetic cells being developed. These include the use of the plastid encoded 16S rRNA gene using primers biased towards marine phototrophs (Fuller *et al.*, 2006a; Lepère *et al.*, 2009; Kirkham *et al.*, 2011), the use of functional genes encoding components of the photosynthetic apparatus (Zeidner *et al.*, 2003; Man-Aharonovich *et al.*, 2010), the use of 18S rRNA gene primers specific to photosynthetic taxa (Viprey *et al.*, 2008) and the construction of clone libraries from flow cytometry sorted populations (Shi *et al.*, 2009, 2011; Marie *et al.*, 2010). These approaches showed that the classes Prasinophyceae, Chrysophyceae and Prymnesiophyceae are important components in open oceans and are more diverse than previously thought.

The Atlantic Ocean is the second largest ocean on the planet, covering an area of approximately 82 million km<sup>2</sup>, incorporating a varied range of trophic regimes. Longhurst (2007) used physical and biological data, from a range of sources, to classify the ocean into a range of biomes, which were further portioned into biogeochemical provinces based on topography, patterns of stratification and seasonality (Figure 4.1). The biomes sampled on the AMT cruises were the Atlantic coastal biome (Falkland current convergence zone (FKLD)), Atlantic west wind biome (North Atlantic Drift (NADR), North Atlantic Subtropical Gyre East and West (NAST-E + NAST-W)), Atlantic trade wind biome (North Atlantic Tropical Gyre (NATR), Western Tropical Atlantic (WTRA) and South Atlantic Gyre (SATL)) and the South Subtropical Convergence (SSTC) in the sub polar south west Atlantic (Aiken *et al.*, 2009).

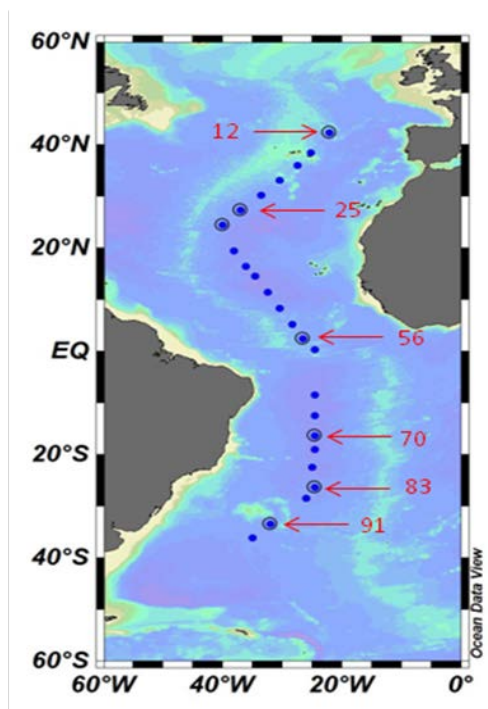


allows for the passage through a range of ecosystems from temperate to tropical and from upwelling systems to eutrophic shelf systems and oligotrophic gyre systems (Figure 4.2).

The AMT was started with the aim of studying phytoplankton processes, bio optical signatures and zooplankton dynamics from the mesoscale to the basin scale.

AMT18, in October and November 2008, was the start of the third phase of the program which is funded by the NERC Oceans 2025 initiative and organised by the Plymouth Marine Laboratory in collaboration with the National Oceanography Centre (Southampton). The project has three specific aims:

- (1) To investigate the changes in the structure, functional properties and trophic status of the plankton, as they vary in time and space.
- (2) To assess how physical processes control the rate of nutrient supply to the planktonic ecosystem.
- (3) To ascertain how photo-degradation and atmosphere-ocean exchange affects the formation and fate of organic matter.



**Figure 4.2. AMT18 transect during October/November 2008. Stations are depicted as blue dots with arrows depicting samples analysed during this study.**

Despite the importance of the eukaryotic fraction of the picophytoplankton to carbon fixation (Li, 1994; Jardillier *et al.*, 2010; Cuvelier *et al.*, 2010), relatively little is understood about the diversity and distribution of these organisms. Overall, the aim of my work was to investigate the diversity of these PPEs at the ocean basin scale and to attempt to ascertain the factors driving the distribution of these organisms in the environment.

## **4.2 Methods**

### **4.2.1 Cruise details**

Samples were obtained during the Atlantic Meridional Transect cruise 18 between 4<sup>th</sup> October-10<sup>th</sup> November 2008. The cruise was undertaken on the British Antarctic Survey ship, the RRS James Clark Ross. The cruise departed from Immingham, UK and headed south towards Port Stanley, Falkland Islands, encompassing northern and



southern gyre and equatorial waters en route. Samples were obtained from six discrete depths (surface, 55%, 33%, 14%, 1% and 0.1% surface light intensity) at 21 stations along the cruise track (Figure 4.2) with a rosette of 20 l Niskin bottles. Light conditions correspond to the total incident PAR, measured using a Chelsea PAR Sensor (UWIRR). Light depths were determined from the previous days noon cast. Temperature was measured using a SBE 3P temperature sensor and a digiquartz temperature compensated pressure sensor. Micromolar nitrate, nitrite, silicate and phosphate concentrations were determined using a five-channel Bran and Luebbe segmented flow colorimetric autoanalyser as described previously (Zwirgmaier *et al.*, 2007). Nutrient concentrations mentioned in the text are values from seawater samples collected from surface waters down to 0.1% light levels unless otherwise stated.

#### **4.2.2 Enumeration of PPEs during the Atlantic Meridional Transect**

PPE numbers were enumerated during AMT18 based on autofluorescence and DNA staining as described in Zubkov *et al.*, (2007).

#### **4.2.3 Sample collection for bulk community DNA**

Bulk community DNA was collected at 6 light depths (97, 55, 14, 7, 1 and 0.1%) from 21 stations during the pre dawn CTD using a stainless steel CTD/Rosette system (SeaBird). Up to 12L of seawater was pre-filtered through 100  $\mu\text{m}$  mesh (Millipore) at each depth before being filtered through 10  $\mu\text{m}$  and 5  $\mu\text{m}$  pore-size polycarbonate filters (different from the previous 3  $\mu\text{m}$  classification of picoeukaryotes) and finally onto a 0.2  $\mu\text{m}$  pore size Supor filter. DNA was subsequently extracted from 4 stations using the protocol described in section 2.2.1.4 (Table 4.1).

**Table 4.1. Percentage light levels and the corresponding depths at CTDs along the transect.**

	CTD			
	12	56	83	91
Light level (%)	42.40°N 22.11°W	2.47°N 25.51°W	26.33°S 24.59°W	33.46°N 32.00°W
97	Surface	-	Surface	Surface
55	12m	10m	17m	9m
14	22m	34m	55m	30m
7	40m	46m	-	41m
1	95m	79m	128m	71m
0.1	-	-	-	100m

#### 4.2.4 Collection of samples for flow cytometric sorting of picoeukaryotes

Seawater samples were collected from two depths (55% light level and DCM) at 6 stations in the oligotrophic northern and southern gyres, the equatorial region and high latitude extremities of the cruise (Table 4.2) in order to target picoeukaryotes by flow cytometry.

**Table 4.2. The depths of CTDs corresponding to the 55% light level and DCM**

	CTD					
	12	25	56	70	83	91
55%	12m	19m	10m	18m	17m	9m
DCM	95m	120m	70m	150m	115m	41m

Seawater concentrates were collected in Cell Traps™ (0.22 µm) after size fractionation (100 µm mesh, 10 µm and 5 µm polycarbonate filter membranes) as described in section 2.3.1.3.

#### 4.2.5 Flow cytometric sorting of picoeukaryotes for nucleic acid extraction

Flow cytometry was used to target PPEs using a MoFlo (Dako-Cytomation) flow cytometer. The cells were stained using SYBR Green 1 and sorted using their chlorophyll *a* autofluorescence (FL3), stained DNA fluorescence (FL1) and side scatter (SSC) (Zubkov *et al.*, 2007). About 50,000 cells were sorted per sample from two PPE size fractions (Euk A and Euk B in Figure 3.2) and stored at -80°C. Nucleic acid was extracted as described in 2.2.1.3.

#### **4.2.6 DNA extraction from filters and sorted cells**

Nucleic acids from flow cytometrically sorted cells were extracted using a modified protocol of Qiagen's DNeasy Blood & Tissue kit as described in section 2.2.1.3.

Nucleic acids from total phytoplankton filters were extracted using a phenol chloroform approach detailed in section 2.2.1.4.

#### **4.2.7 PCR amplification of 16S and 18S rRNA genes**

PPEs were specifically targeted using the plastid biased 16S rRNA primers PLA491F and OXY1313R (Fuller *et al.*, 2006a) using the PCR conditions as detailed in section 2.2.2.1.

Nuclear 18S rRNA genes were amplified using the primers 82F and 1498R (Lopez Garcia *et al.*, 2003) using the PCR conditions as described in section 2.2.2.2. Prymnesiophytes were specifically targeted using specific primers (Jardiller, unpublished) using PCR products from the nuclear 18S rRNA amplification as templates in a nested PCR approach (section 2.2.2.3)

#### **4.2.8 Construction of clone libraries**

Purified PCR products (section 2.2.3) were used to create clone libraries using the pGEM® T Easy Vector system (Promega) as detailed in section 2.2.4.2. Positive clones were selected and 1 µl was used for subsequent PCR amplification (section 4.2.6). The resulting PCR products were analysed using restriction endonuclease digestion. A double digest (*HaeIII* and *EcoRI*) was used for the plastid biased 16S rRNA PCR products whilst a single digest (*HaeIII*) was used for the nuclear 18S rRNA PCR products (section 2.2.5). PCR products of clones with unique restriction fragment length polymorphisms (RFLP) patterns, were sent to the NERC Biomolecular Analysis Facility in Edinburgh for Sanger sequencing, using an ABI 3730 instrument.

#### **4.2.9 Phylogenetic analysis**

Forward and reverse reads were assembled using Seqman (DNASTAR) and alignments were manually checked to remove any ambiguous bases. Resulting assemblies were checked for chimeras as detailed in section 2.4.2.2. Phylogenetic analysis was undertaken using a laboratory curated ARB database as described in 2.4.2.3.

#### **4.2.10 Statistical analysis**

Coverage values of the clone libraries were calculated as described in section 2.4.1.2. Species richness was estimated using the non parametric method Chao1 (Colwell and Coddington, 1994) as detailed in 2.4.1.1. Species diversity was calculated using the Margalef diversity index as described in 2.4.1.3 with species composition of the clone libraries being calculated using the Morisita Horn equation (section 2.4.1.4). To explain the distribution of PPEs canonical correspondence analysis was undertaken as described in section 2.4.1.6.

## **4.3 Results**

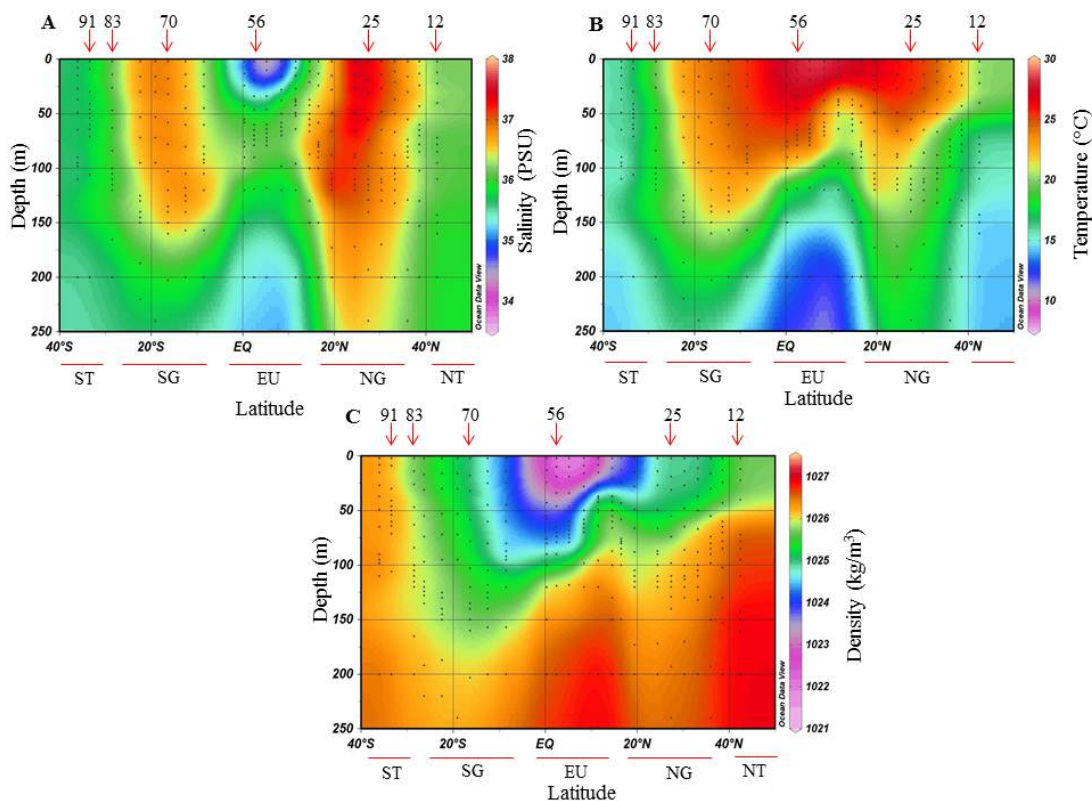
### **4.3.1 Ancillary measurements taken during AMT18**

Direct measurements of temperature (Figure 4.3A) and salinity (Figure 4.3B) were measured during the CTD casts undertaken daily on the AMT cruise.

The highest temperatures observed during AMT18 were observed in the equatorial region between 0° and 10°N. A temperature of 28.7°C was measured at 8°N at a depth of 2m. The northern and southern gyre regions were characterised by temperatures ~ 25°C. In the southern gyre thermal stratification occurred to a depth of ~ 100m and to 75m in the northern gyre. The temperate regions of the cruise were characterised by lower temperatures. The northern temperate region (> 40°N) had water temperatures ~ 20°C. The southern temperate region (> 30°S) had lower temperatures (~ 15°C) due to the cruise occurring during the southern hemisphere spring.

Two regions of high salinity (~ 37 PSU) were observed along AMT18 corresponding to the two gyre regions. The lowest salinity values measured (34 PSU) were observed in the equatorial upwelling region. Both the northern and southern temperate regions were characterised by salinity values around 36 PSU.

Density values were lowest in the equatorial upwelling region reaching values of 1021 kg/m<sup>3</sup> at a depth of 10m in station 51.



**Figure 4.3. A) Salinity (PSU), B) Temperature (°C) and C) density (kg/m<sup>3</sup>) along the AMT18 cruise track. Black dots represent sampling points and red arrows depict the stations used for this study. Red bars denote different regions along the AMT cruise. NT = Northern temperate, NG = Northern gyre, EU = Equatorial upwelling, SG = Southern gyre and ST = Southern temperate.**

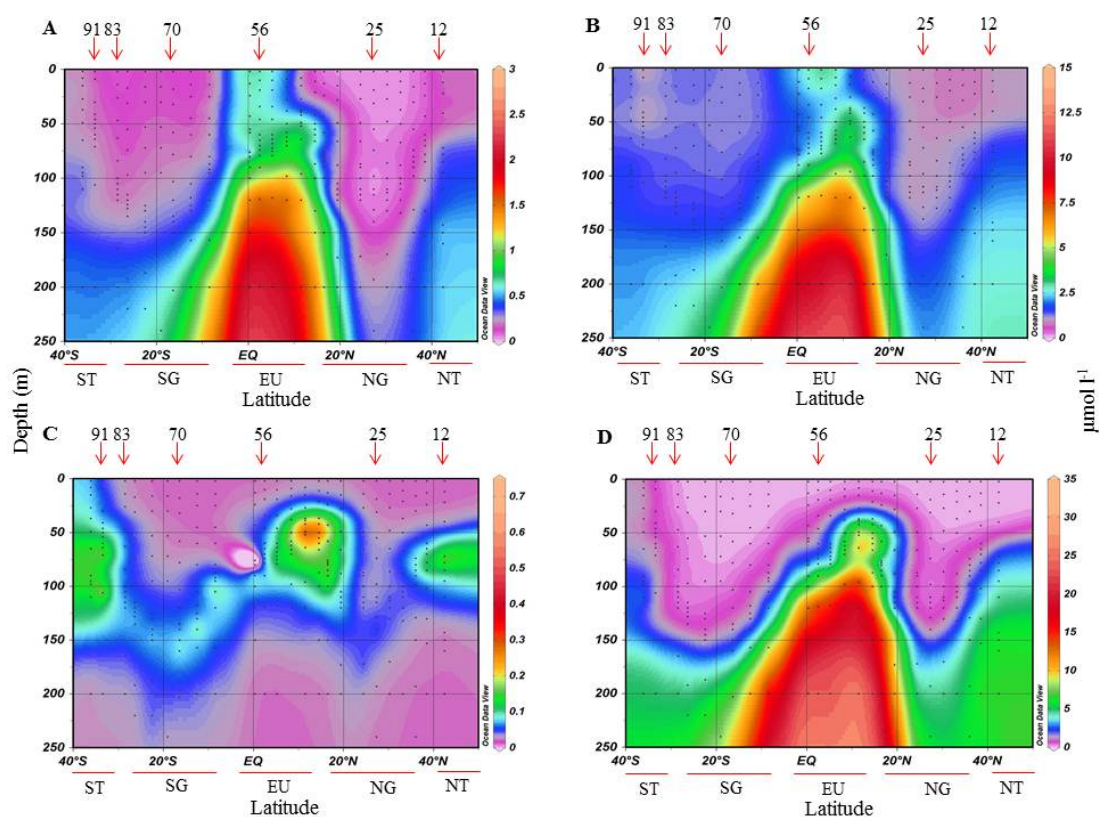
Nutrient concentrations, determined as in section 4.2.1, were routinely measured throughout the cruise by Malcolm Woodward and Carolyn Harris from Plymouth Marine Laboratory. Their data is used here to produce ODV plots depicting the variation in nutrient concentrations with depth and latitude (Figure 4.4).

Phosphate levels were less than  $0.1 \mu\text{mol l}^{-1}$  across a large proportion of the AMT (Figure 4.4A). The level of phosphate depletion was greatest in the northern gyre with values of phosphate  $<0.1 \mu\text{mol l}^{-1}$  to depths of 150m. A region of higher phosphate concentration ( $>0.5 \mu\text{mol l}^{-1}$ ) was measured in the equatorial upwelling region.

The distribution of silicate (Figure 4.4B) followed a similar pattern to that of phosphate. The lowest concentrations of silicate were observed in the northern gyre region ( $2.3 \mu\text{mol l}^{-1}$ ). In the equatorial upwelling region increased concentrations of silicate were observed in surface waters, up to  $3.5 \mu\text{mol l}^{-1}$  at station 51 ( $5.2^\circ\text{N}$   $28.31^\circ\text{W}$ ).

Surface waters along the AMT were characterised by low levels of nitrite (Figure 4.4C). Peaks in nitrite concentrations were measured in sub surface waters. Increased levels of nitrite were observed at greater depths in the gyre regions than in the temperate and upwelling regions. The peak in nitrite concentration ( $0.6 \mu\text{mol l}^{-1}$ ) was measured at 50m at  $14.5^\circ\text{N}$  in the equatorial upwelling region.

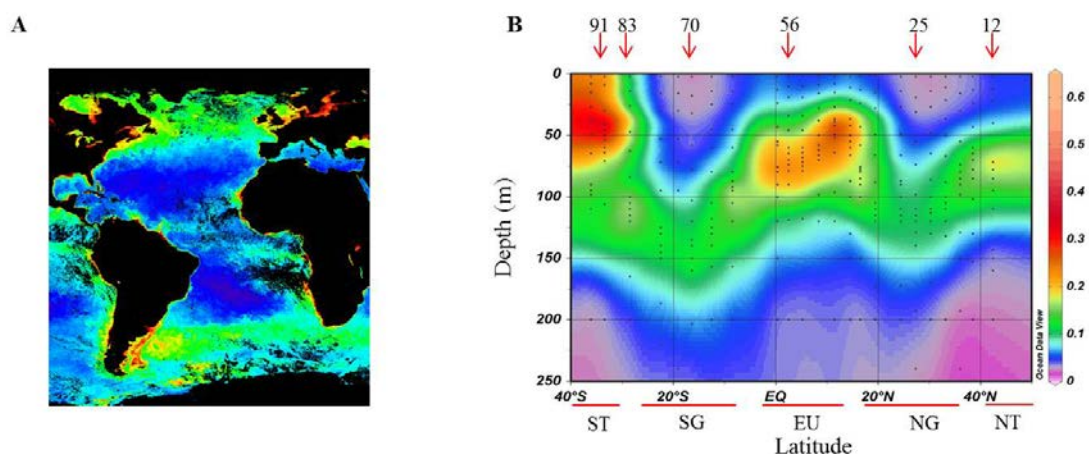
Measurements of combined nitrate + nitrite concentrations were also undertaken during the cruise. Nitrate values were calculated by the subtraction of the corresponding nitrite measurements (Figure 4.4D). In surface waters, along the whole of the AMT, levels of nitrate were extremely low ( $< 0.1 \mu\text{mol l}^{-1}$ ). Increases of nitrate were observed in the sub surface in the equatorial upwelling region.



**Figure 4.4. Nutrient profiles along AMT18. A) Phosphate ( $\mu\text{mol l}^{-1}$ ); B) Silicate ( $\mu\text{mol l}^{-1}$ ); C) Nitrite ( $\mu\text{mol l}^{-1}$ ); D) Nitrate ( $\mu\text{mol l}^{-1}$ ). Black dots represent sampling points and red arrows depict the stations used for this study. Red bars denote different regions along the AMT cruise. NT = Northern temperate, NG = Northern gyre, EU = Equatorial upwelling, SG = Southern gyre and ST = Southern temperate.**

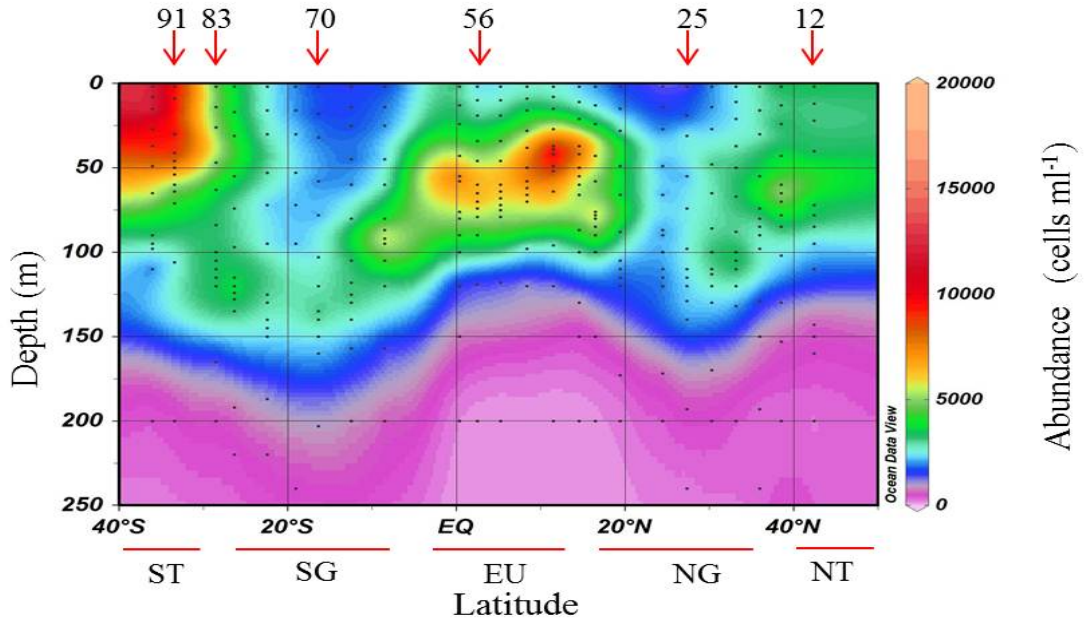
Chlorophyll *a* levels during AMT18 were estimated using the MODIS satellite system and a composite for October 2008 was generated (Figure 4.5A). Depth profiles of chlorophyll *a* were also measured during the CTD casts (Figure 4.5B). The gyre regions exhibited low concentrations of chlorophyll *a* in surface waters. Peaks in chlorophyll *a* concentrations were observed in subsurface regions designated as the DCM. The highest chlorophyll *a* concentrations were observed in the southern temperate region reaching  $0.49 \text{ mg m}^{-3}$  at 41m at  $33.46^\circ\text{S}$ .





**Figure 4.5. Chlorophyll *a* content along the AMT18.** A) Composite of chlorophyll *a* content from MODIS satellite system during October 2008; B) Measured chlorophyll *a* content from CTD casts during AMT18. Black dots represent sampling points and red arrows depict the stations used for this study. Red bars denote different regions along the AMT cruise. NT = Northern temperate, NG = Northern gyre, EU = Equatorial upwelling, SG = Southern gyre and ST = Southern temperate.

Analyses of picophytoplankton cell abundances by flow cytometry were undertaken by Ross Holland (National Oceanographic Centre, Southampton) during the cruise (see section 4.2.2). The highest concentrations of PPEs were observed  $\sim 40^{\circ}\text{S}$  with values reaching approximately  $19,000 \text{ cells ml}^{-1}$  (Figure 4.6). In the gyre regions concentrations of PPEs were much lower, averaging about  $1000 \text{ cells ml}^{-1}$  in surface waters. Increased abundances of PPEs were observed in subsurface waters, corresponding to the DCM. Higher concentrations of picoeukaryotes were observed in the equatorial upwelling region compared to the gyres.

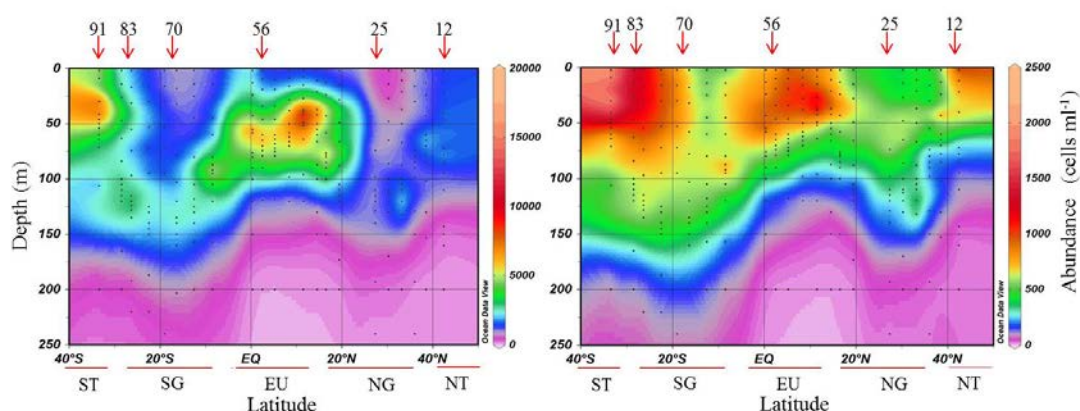


**Figure 4.6. Abundance of PPEs (cells ml<sup>-1</sup>) along AMT18. Black dots represent sampling points and red arrows depict the stations used for this study. Red bars denote different regions along the AMT cruise. NT = Northern temperate, NG = Northern gyre, EU = Equatorial upwelling, SG = Southern gyre and ST = Southern temperate.**

As well as assessing total PPE abundance along AMT18, the abundance of specific PPE populations, i.e. Euk-A and Euk-B (see Jardillier *et al.*, 2010; section 4.2.4) was also determined, so that subsequent diversity studies performed here could be mapped onto <sup>14</sup>C CO<sub>2</sub> fixation data for these groups (Grob *et al.*, in press). The Euk-A PPE population had peak abundances in the subsurface throughout the transect (Figure 4.7). In contrast, the gyre regions were characterised by low concentrations of the Euk-A population. Peak abundances of Euk-A were observed in the equatorial upwelling region with concentrations reaching 17,000 cells ml<sup>-1</sup>.

Throughout the transect Euk-B cell abundances were lower than those of Euk-A. Low concentrations (~500 cells ml<sup>-1</sup>) were observed in the gyre regions. Peak

abundances of Euk-B were observed in the southern temperate region reaching 2484 cells  $\text{ml}^{-1}$  at 30m depth at 33.46°S.



**Figure 4.7. Flow cytometry based cell abundance measurements of specific PPE populations along AMT18. A) Abundance of Euk-A (cells  $\text{ml}^{-1}$ ); B) Abundance of Euk-B (cells  $\text{ml}^{-1}$ ). Black dots represent sampling points and red arrows depict the stations used for this study. Red bars denote different regions along the AMT cruise. NT = Northern temperate, NG = Northern gyre, EU = Equatorial upwelling, SG = Southern gyre and ST = Southern temperate.**

#### 4.3.2 Taxonomic composition of total PPEs from bulk community filters

To assess total PPE diversity along both the latitudinal gradient encompassed by AMT18 and down a depth profile, plastid biased 16S rRNA primers were used on community bulk DNA filters. DNA was extracted from four stations (Figure 4.2) and for each station at six depths (Table 4.1). 16S rRNA gene amplicons produced from the PCR using plastid-biased PCR primers (see section 2.2.2) were cloned into the pGEM cloning vector (see section 2.2.4) and PCR products from individual clones subsequently screened by RFLP analysis (see section 2.2.5). Unique RFLP patterns were completely sequenced bidirectionally using standard procedures (see section 2.2.6) and OTUs defined by relationships in phylogenetic trees.

For each of the 19 clone libraries constructed the number of clones obtained per library, the number of OTUs identified, the number of sequences per library, and the coverage value for each library were determined (see Table 4.3), the latter indicating how well the diversity within each sample was estimated by the library.

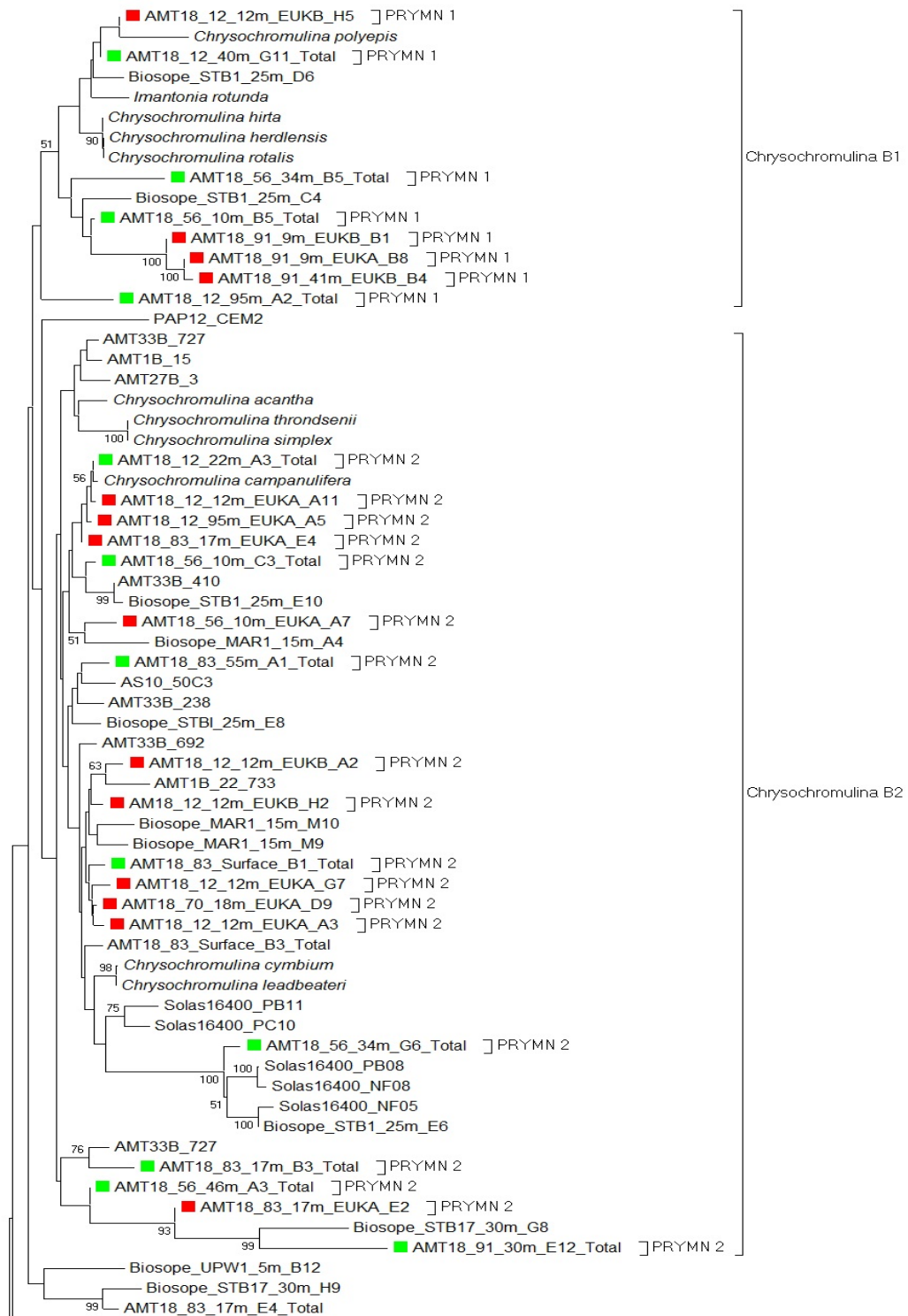
**Table 4.3. Statistics of clone libraries created from bulk community filters**

Station	Depth (m)	Clones (RFLPs)	OTUs	Coverage	Margalef index
12 Northern Temperate	Surface	57 (4)	4	1	0.74
	12	20 (6)	5	1	1.34
	22	40 (4)	4	0.95	0.81
	40	49 (5)	5	0.96	1.02
	95	41 (4)	4	0.98	0.81
56 Equatorial Upwelling	10	25 (5)	4	0.96	0.93
	34	24 (8)	7	0.91	1.89
	46	16 (3)	3	0.94	0.72
	79	14 (5)	5	0.71	1.52
83 Southern Gyre	Surface	38 (4)	3	1	0.54
	17m	37 (6)	5	0.97	1.11
	55m	44 (5)	4	0.97	0.79
	128m	19 (6)	5	1	1.36
91 Southern Temperate	Surface	82 (6)	5	0.99	0.91
	9	45 (7)	6	0.98	1.31
	30	68 (6)	6	1	1.19
	41	40 (6)	6	0.93	1.36
	71	69 (6)	4	1	0.71
	100	37 (4)	4	0.97	0.83

Prymnesiophyceae were observed to be the dominant class throughout AMT18 in these bulk DNA clone libraries. In the northern temperate region (CTD12) the prymnesiophyte component was dominated by sequences related to *Chrysochromulina* clade B2 (Prymn 2) to a depth of 22m and *Chrysochromulina* clade B1 (Prymn 1) at depths of 40m and 95m. At all depths sequences related to Prym 16S-II (Prymn 6) were observed (see Figure 4.8). A third OTU was observed at 40m with sequences clustering in the order Coccolithales (Prymn 3). In the

equatorial upwelling region (CTD 56) 3 OTUs relating to prymnesiophytes were defined. The dominant OTU found at all depths was related to *Chrysochromulina* clade B2 (Prymn 2). Sequences related to *Chrysochromulina* clade B1 (Prymn 1) (within the order Prymnesiales) and an environmental clade (Prymn 4) were also obtained at this station (see Figure 4.8). In the southern gyre (CTD 83) one OTU assigned to *Chrysochromulina* clade B2 (Prymn 2) was found at all depths. In addition, an OTU related to Coccolithales (Prymn 3) was found at all depths. At 17m an OTU related to Phaeocystales (Prymn 5) was observed, while sequences clustering in Prym 16S-II (Prymn 6) were found at 128m. In the southern temperate region (CTD 91) 2 prymnesiophyte OTUs were observed at all depths. These OTUs were related to Phaeocystales (Prymn 5) and Prym 16S-II (Prymn 6). A third OTU was present at all depths except 71m with sequences clustering in the order Coccolithales (Prymn 3) (Figure 4.8).

Members of the Prasinophyceae were minor components of all the clone libraries constructed (Figure 4.9). In the northern temperate region (CTD 12) one OTU (Pras 1) was observed at 95m which was phylogenetically affiliated to clade Pras 16S-VIII. In the equatorial upwelling region (CTD 56) 2 OTUs were observed (Figure 4.10). One OTU (Pras 3) observed at 34m, was related to *Pyramimonas* while the second OTU from 46m depth, was related to prasinophyte clade VIIA (Pras 4). Prasinophyceae were present in low numbers at all depths except the surface in the southern temperate region (CTD 91). The prasinophyte component was comprised of one OTU (Pras 1) related to Pras 16S-VIII, a clade comprising only environmental sequences and lacking any cultured isolates (see Lepère *et al.*, 2009 for designation of this clade).



**Figure 4.8. Phylogenetic tree of Prymnesiophyceae sequences with AMT18 sequences showing corresponding OTU numbers. Clones created from total DNA filters are marked with a green box whilst those from flow cytometrically sorted cells are marked with a red box.**



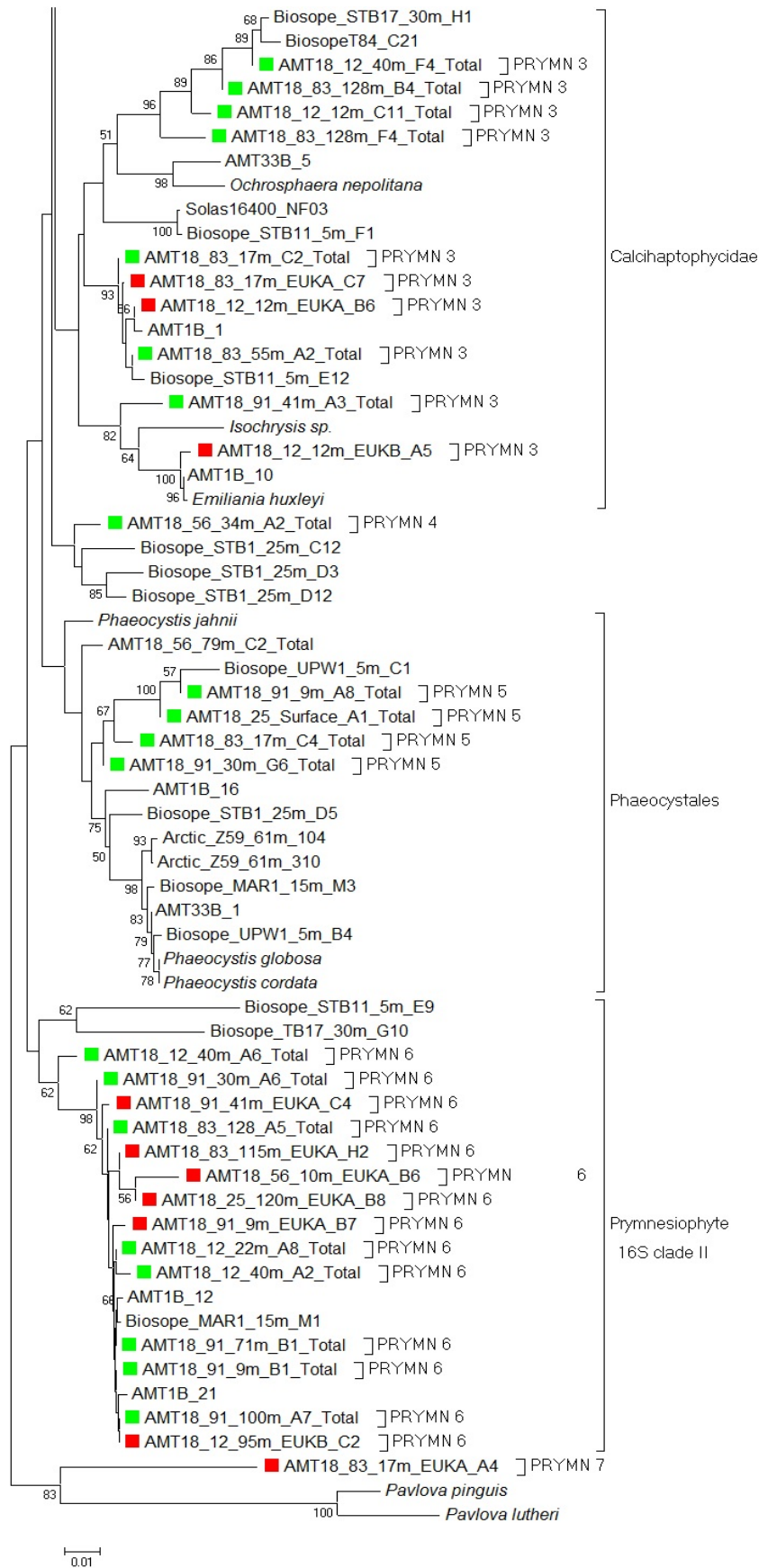
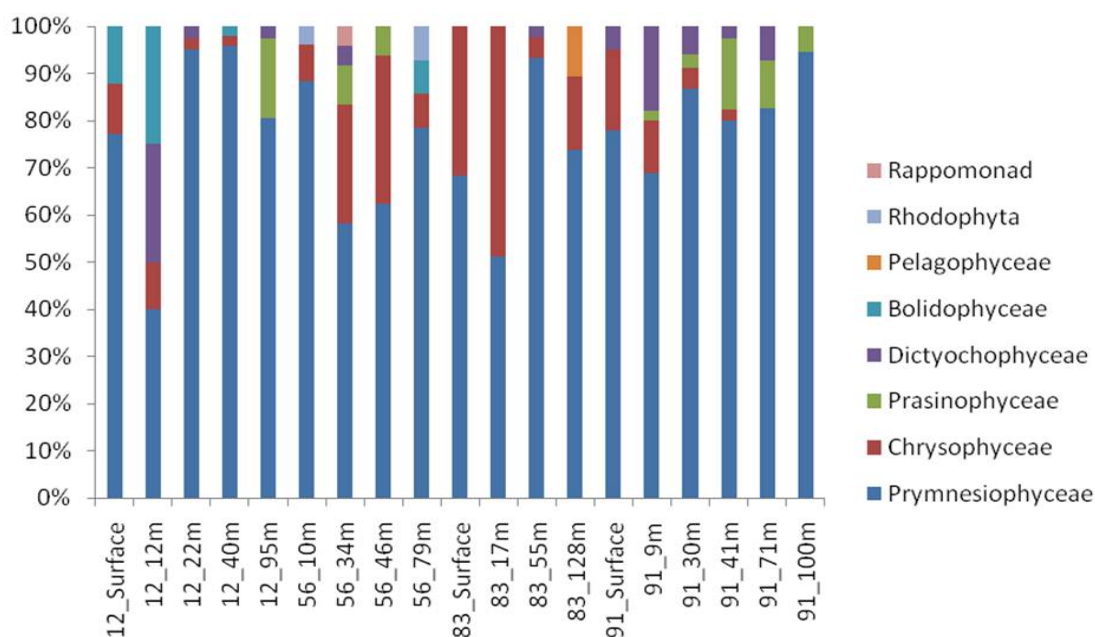


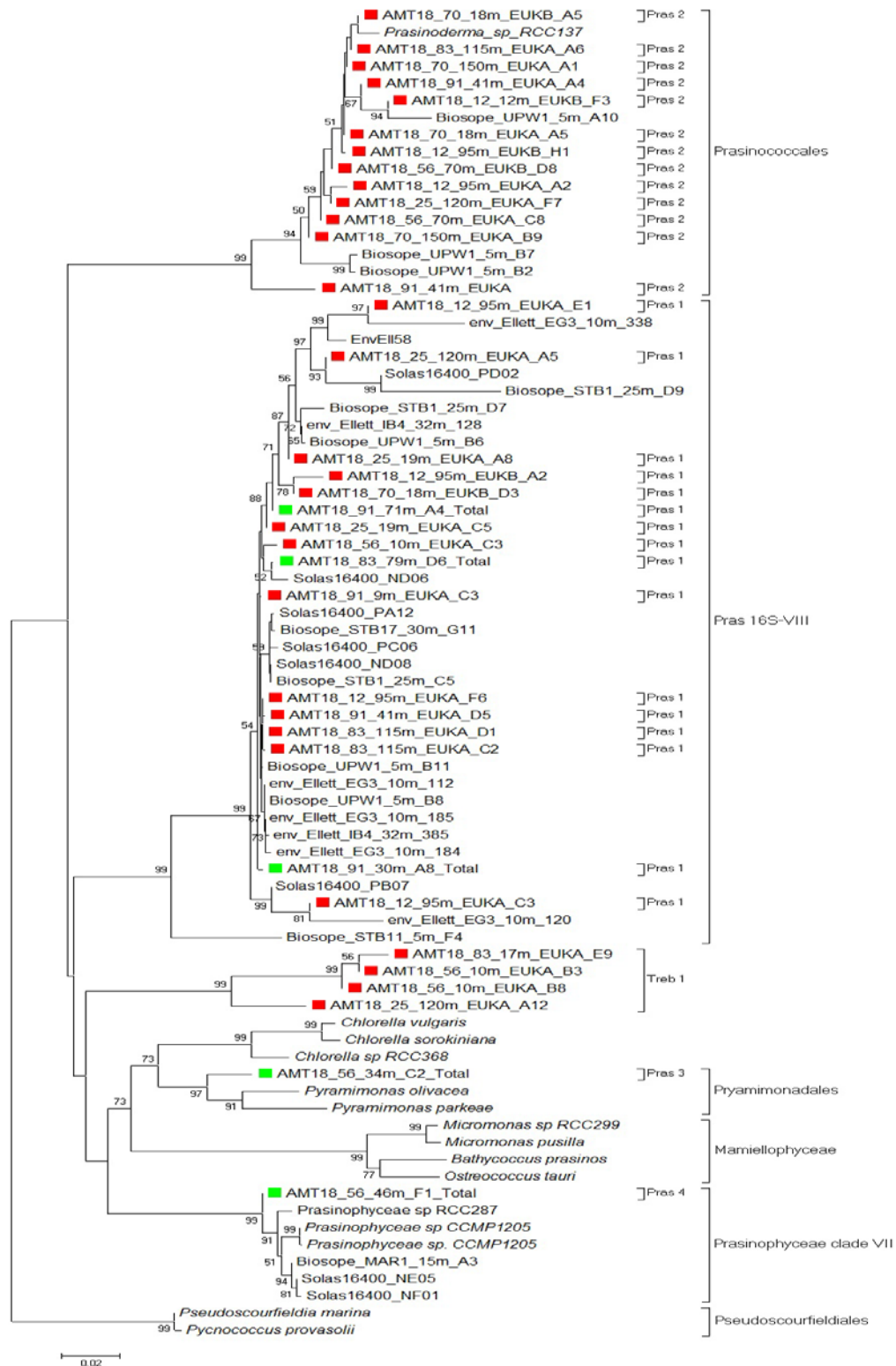
Figure 4.8 cont. Phylogenetic tree of Prymnesiophyceae sequences.



**Figure 4.9. The taxonomic composition of PPEs along AMT18 using plastid biased 16S rRNA primers and bulk community DNA.**

Sequences phylogenetically affiliated to the class Chrysophyceae were also found at all stations. In the northern temperate region chrysophytes were observed in the CTD12 clone libraries to a depth of 40m. The percentage contribution of chrysophytes to the clone libraries declined with depth (see Figure 4.9). This observation was also observed in the southern temperate region (CTD 91) where sequences related to chrysophytes declined from 17% at the surface to 2.5% at 41m. In the equatorial upwelling region (CTD 56) the contribution of chrysophytes to the clone libraries was highest at 46m accounting for 31% of the clone library. In the southern gyre (CTD 83) chrysophytes made a significant contribution to the clone libraries at both the surface and 17m with smaller contributions at 55m and 128m. The Chrysophyceae sequences obtained in these libraries were only loosely related to cultured representatives but clustered with environmental sequences from previous AMT cruises (Figure 4.11).



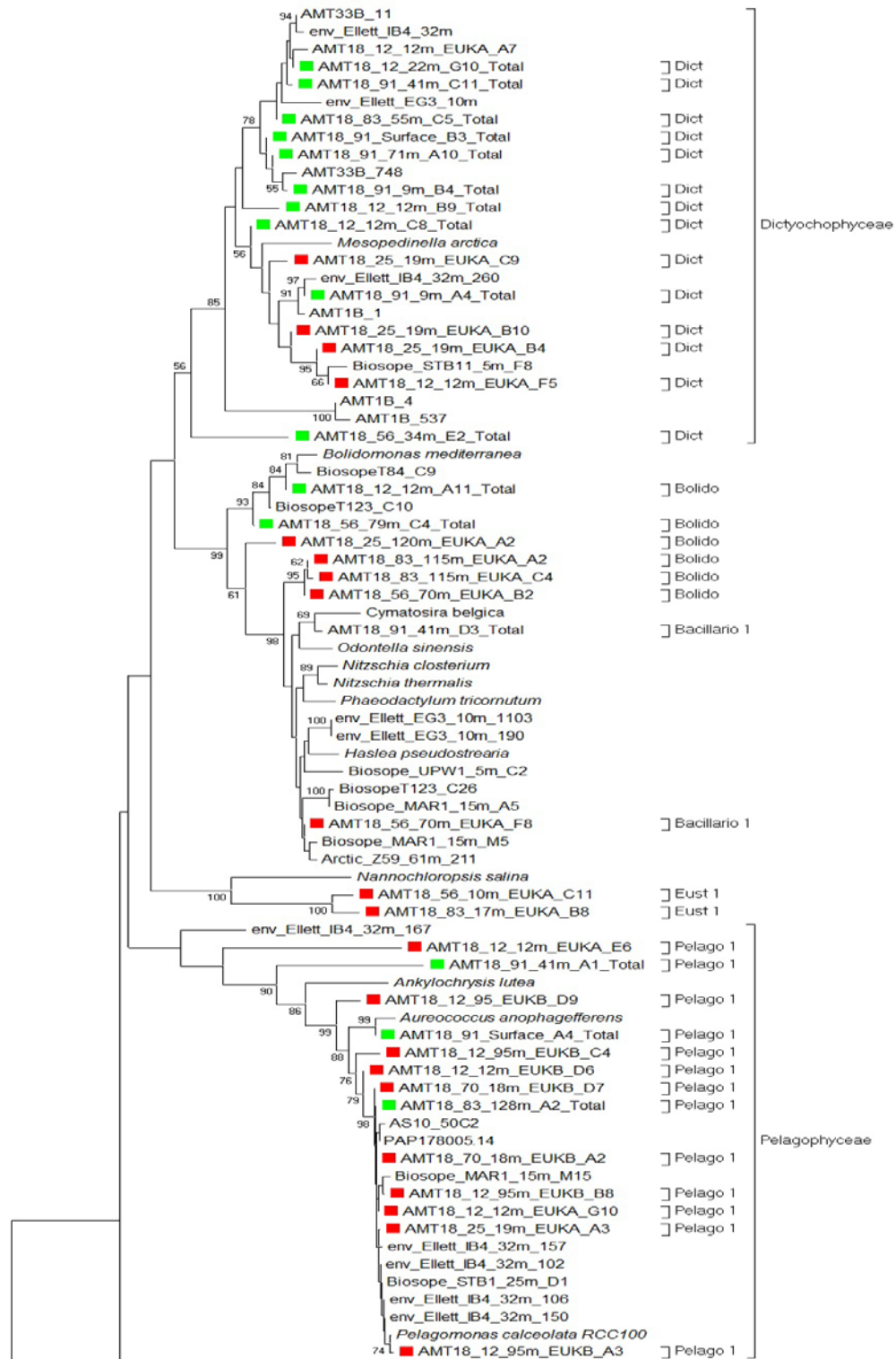


**Figure 4.10. Phylogenetic tree of Chlorophyta sequences with AMT18 sequences showing corresponding OTUs. Red boxes indicate sequences obtained from flow cytometrically sorted cells whilst green boxes indicate clones from total DNA filters.**

Dictyochophyceae were present at all stations in low numbers. In the northern temperate region (CTD 12) Dictyochophyceae were present at 12m, accounting for 25% of the library. Dictyochophyceae were observed in low numbers at 22m and 95m (Figure 4.9). In the equatorial upwelling (CTD 56) Dictyochophyceae were observed at 34m. One OTU was also observed in the southern gyre (CTD 83) at 55m. In the southern temperate region (CTD 91) Dictyochophytes were present in small numbers to depths of 71m.

Small contributions from Bolidiophyceae, rappemonads and Rhodophyta were observed in the equatorial upwelling region (CTD 56). Bolidiophyceae were also present in the northern temperate region (CTD 12) accounting for 25% of the clone library at 12m (see Figure 4.9).

The composition of the libraries down the depth profile was compared using the Morisita Horn Similarity Index. In the northern temperate region the composition of the clone libraries at 40m and 95m was different to those libraries constructed higher in the water column. The difference in composition is due to the switch from sequences related to *Chrysochromulina* clade B2 to *Chrysochromulina* clade B1 deeper in the water column. In the equatorial upwelling region (CTD 56) the clone library at 34m had a lower similarity compared with the other depths. Coverage values for the libraries in the equatorial upwelling were low due to the amplification of sequences related to *Prochlorococcus* dominating the libraries. In the southern temperate region (CTD 91) there was a high similarity between all depths and with a high coverage value ( $> 0.859$  at all depths).



**Figure 4.11. Phylogenetic tree of Heterokonta sequences with AMT18 sequences showing corresponding OTUs. Red boxes indicate sequences obtained from flow cytometrically sorted cells whilst green boxes indicate clones from total DNA filters.**

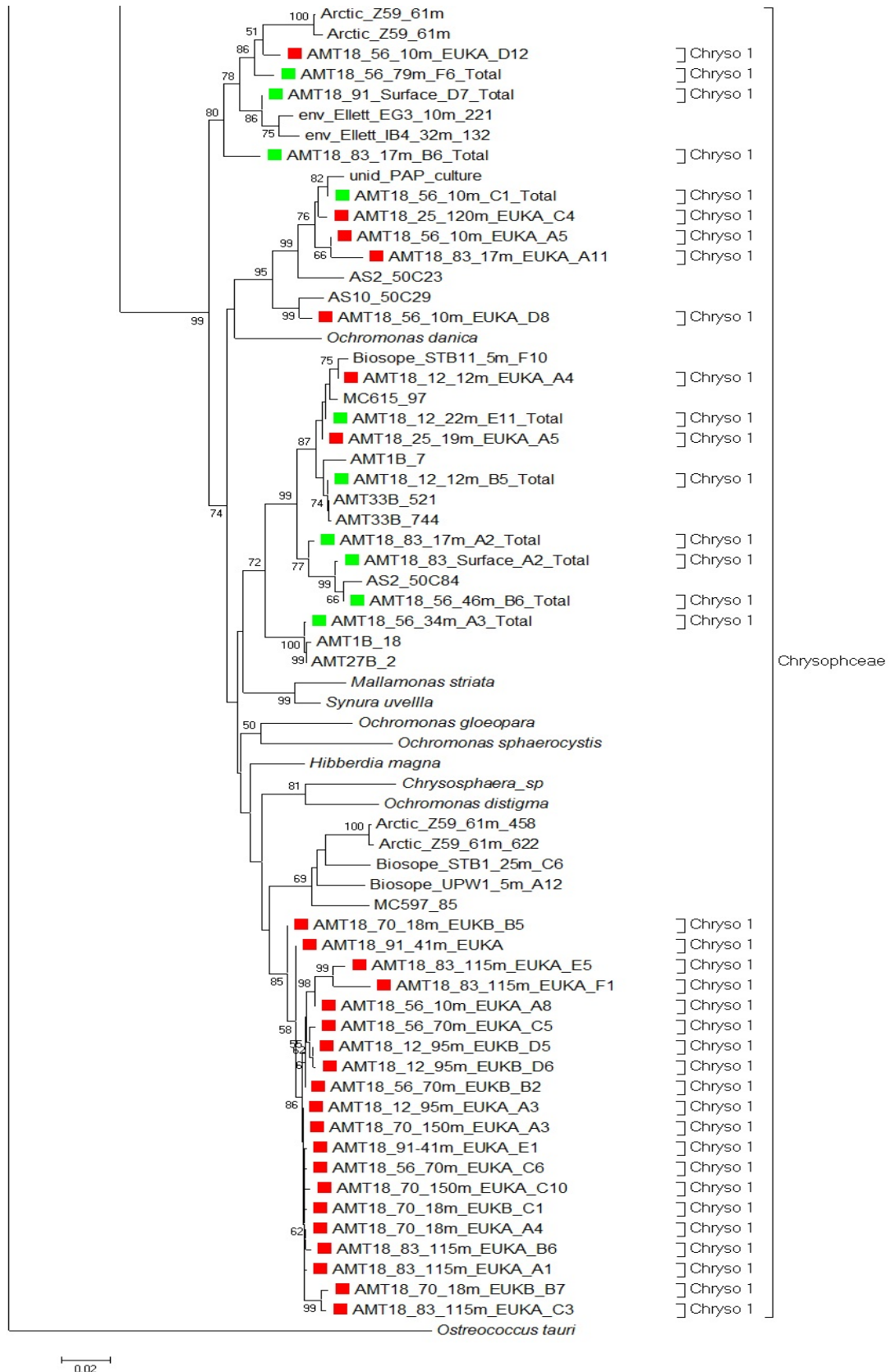


Figure 4.11 cont.

### 4.3.3 Taxonomic diversity of Euk-A and Euk-B PPE populations along AMT18

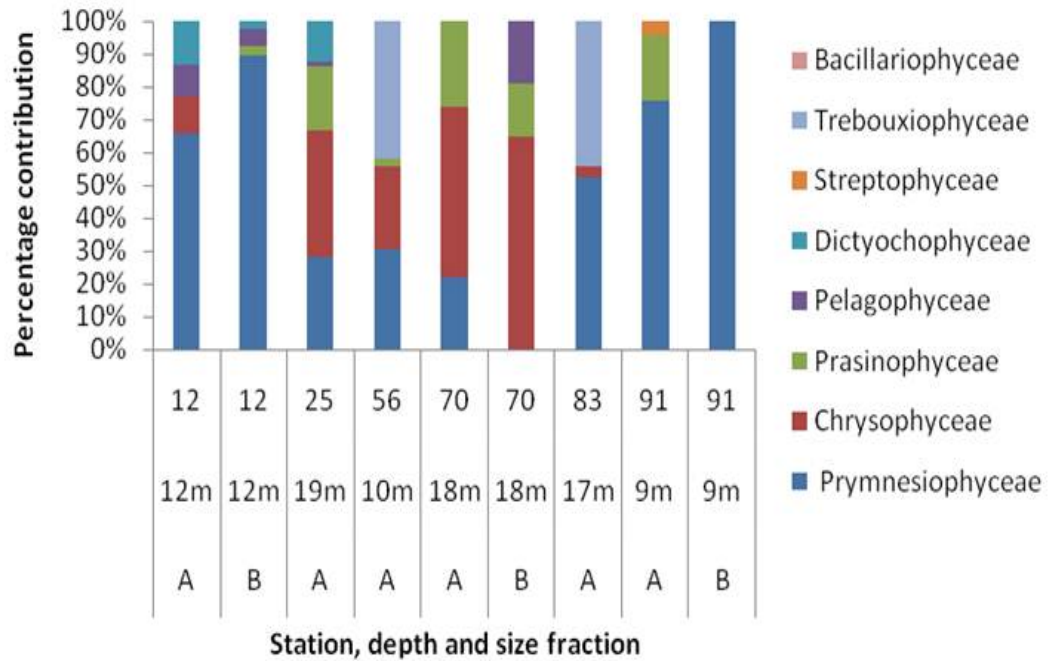
Investigation of the structure of Euk-A and Euk-B PPE populations along AMT18 was performed following flow cytometric sorting of these specific population types using criteria defined in section 4.2.4. Subsequent PCR, of DNA extracted from such populations (see section 2.2.1.3), using marine algal plastid biased 16S rRNA gene primers (Fuller *et al.*, 2006a) was used to construct clone libraries of Euk-A and Euk-B PPE populations at several stations along AMT18 (Figure 4.2; Table 4.2) either in at the 55% light level or at the DCM.

For each of the 18 clone libraries constructed the number of clones obtained per library, the number of OTUs identified, the number of sequences per library, and the coverage value for each library was determined (see Table 4.4). Estimates of coverage ranged from 0.92 to 1, suggesting the majority of predicted OTUs in the environment had been identified with the sampling effort undertaken. This was reinforced with the non parametric test, Chao 1, where the number of observed OTUs was similar to the values predicted by the test.

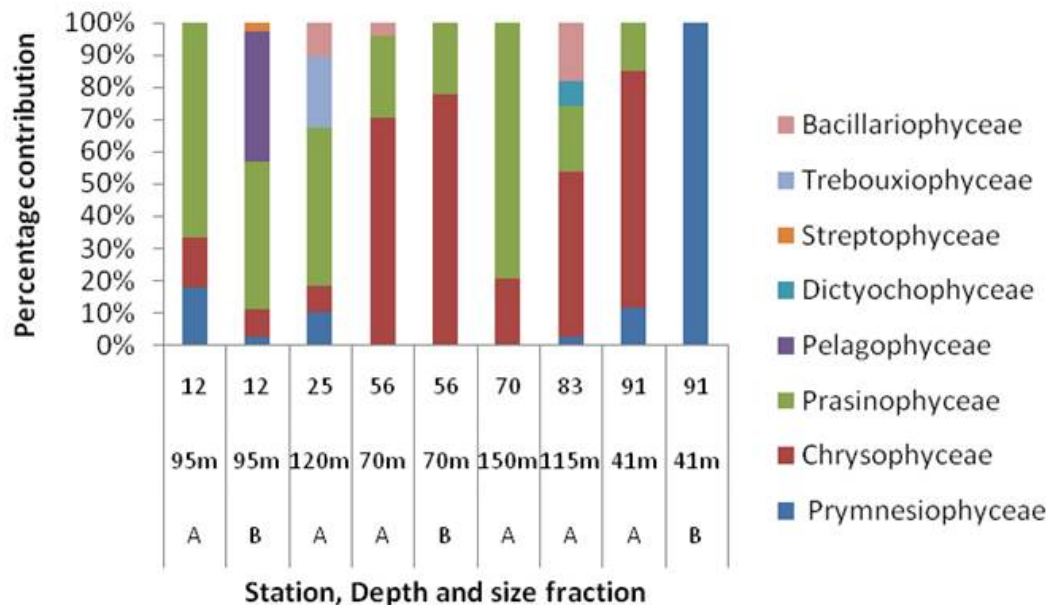
**Table 4.4 Clone library details from flow sorted Euk-A and Euk-B populations along AMT18 including sampling effort and Margalef species richness estimates.**

CTD	Depth	PPE population	Clones (RFLPs)	OTUs	Chao1(SD)	Coverage	Species Richness
12	12	Euk-A	61 (7)	5	5 (0.44)	0.98	0.97
		Euk-B	37 (7)	7	7 (0.92)	0.94	1.66
	95	Euk-A	39 (7)	6	6 (0.09)	1	1.36
		Euk-B	35 (10)	6	6 (1.29)	0.94	1.40
25	19	Euk-A	57 (7)	6	6 (0.25)	0.98	1.24
	120	Euk-A	49 (7)	7	8 (2.23)	0.96	1.54
56	10	Euk-A	37 (8)	6	9 (4.41)	0.92	1.38
	70	Euk-A	213 (4)	4	4 (0)	1	0.56
		Euk-B	159 (2)	2	2 (0)	1	0.20
70	18	Euk-A	46 (5)	4	4 (0)	1	0.78
		Euk-B	37 (6)	3	3 (0)	1	0.55
	150	Euk-A	45 (4)	2	2 (0)	1	0.26
83	17	Euk-A	64 (7)	6	6 (0.01)	1	1.20
	115	Euk-A	39 (11)	6	6 (0.25)	0.97	1.36
91	9	Euk-A	45 (5)	5	5 (0.14)	0.98	1.05
		Euk-B	50 (1)	1	-	1	0.00
	41	Euk-A	34 (8)	4	4 (0.43)	0.97	0.85
		Euk-B	50 (1)	1	-	1	0.00

The percentage contribution of specific PPE classes in each of the Euk-A and Euk-B populations analysed was also determined at either the 55% light level (Figure 4.12, Table 4.5) or at the DCM (Figure 4.13, Table 4.5).



**Figure 4.12.** The taxonomic composition of Euk-A and Euk-B PPE populations along AMT18 at the 55% light level using 16S rRNA gene clone libraries and primers targeted to marine algal plastids.



**Figure 4.13.** The taxonomic composition of Euk-A and Euk-B PPE populations along AMT18 at the DCM using 16S rRNA gene clone libraries and primers targeted to marine algal plastids.



**Table 4.5. Percentage contribution of classes to clone libraries from flow cytometrically sorted samples**

Station	Fraction	Depth (m)	Prymn	Chryso	Pras	Pelago	Dict	Treb	Bacill	Strep
12	EUKA	12	66	11	-	10	13	-	-	-
	EUKB		89	-	3	5	3	-	-	-
	EUKA	95	18	15	67	-	-	-	-	-
	EUKB		3	9	46	40	-	-	-	3
25	EUKA	12	28	39	19	2	12	-	-	-
	EUKA	120	10	8	49	-	-	22	10	-
56	EUKA	10	31	25	3	-	-	42	-	-
	EUKA	70	-	70	25	-	-	-	4	-
	EUKB		-	78	22	-	-	-	-	-
70	EUKA	18	22	52	26	-	-	-	-	-
	EUKB		-	65	16	19	-	-	-	-
	EUKA	150	-	21	79	-	-	-	-	-
83	EUKA	17	52	3	-	-	-	44	-	-
	EUKA	115	3	51	21	-	8	-	18	-
91	EUKA	9	76	-	20	-	-	-	-	4
	EUKB		100	-	-	-	-	-	-	-
	EUKA	41	12	74	15	-	-	-	-	-
	EUKB		100	-	-	-	-	-	-	-

In temperate waters Prymnesiophyceae were the dominant class in surface waters i.e. at the 55% light level. In the Euk-A PPE population Prymnesiophyceae accounted for 65% of clones in the northern temperate region and 75% of clones in the southern temperate region. In the Euk-B population Prymnesiophyceae accounted for between 90 and 100% of the clone libraries in the northern and southern temperate regions, respectively. In the northern temperate region (CTD 12) the majority of prymnesiophyte clones were related to *Chrysochromulina* clade B2 (OTU Prymn 2, see Table 4.6, Figure 4.8). One OTU (OTU Prymn 3, Figure 4.8) comprising sequences related to *Ochrosphaera neapolitana* (order Coccolithales) were present in both the Euk-A and B clone libraries (see Table 4.6). In the southern temperate region (CTD 91) the dominant prymnesiophytes in both Euk-A and Euk-B



populations was an OTU (OTU Prymn1) related to *Chrysochromulina* clade B1 (Figure 4.8). Clones related to *Chrysochromulina* clade B2 (OTU Prymn 2) and the Prym 16S-II (OTU Prymn 6) clade were also observed in the Euk-A clone library. The contribution of prymnesiophytes to the clone libraries constructed from the gyre and equatorial upwelling regions was lower than that found in the temperate regions. In the northern gyre (CTD25), Prymnesiophyceae accounted for 28% of the clones in the Euk-A clone library (Figure 4.12) with the majority of these sequences being related to the Prym 16S-II clade (OTU Prymn 6). Sequences related to *Chrysochromulina* clade B2 (OTU Prymn 2) were a minor component in the library. In the equatorial upwelling region (CTD56) sequences related to *Chrysochromulina* clade B2 (OTU Prymn 2) accounted for 33% of the Euk-A clone library. *Chrysochromulina* clade B2 related sequences were present in the southern gyre station (CTD 70) along with sequences clustering in Coccolithales (OTU Prymn 3). Here they accounted for 20% of the Euk-A clone library. No prymnesiophytes were observed in the Euk-B population from this region. In the second southern gyre station (CTD 83) clones related to *Chrysochromulina* clade B2 (OTU Prymn 2) (order Prymnesiales) were again observed (Table 4.6; Figure 4.12). However, the dominant prymnesiophyte at this station comprised sequences clustered in the order Coccolithales (OTU Prymn 3). In total, Prymnesiophyceae clones accounted for almost 50% of the Euk-A clone library at this station.

The class Chrysophyceae showed a complimentary distribution to that of the Prymnesiophyceae. In temperate regions (CTD 12) at the 55% light level Chrysophyceae only made a small contribution to the Euk-A population (see Figure 4.12). Chrysophyceae accounted for 38% of the Euk-A clone library in the northern gyre (CTD 25). In the southern gyre (CTD 70), where the prymnesiophyte

contribution was the lowest, chrysophytes accounted for 52% of the Euk-A clone library. In the second southern gyre station (CTD 83) Chrysophyceae only accounted for a small percentage (~ 3%) of the clone library. The majority of the sequences obtained from OTU Chryso 1 were loosely related to *Ochromonas distigma* although several were related to uncultured chrysophytes obtained previously from the Atlantic and Pacific (Figure 4.11).

Prasinophyceae only made small contributions to the clone libraries at the 55% light levels. The abundance of prasinophyte clones was highest in the southern gyre station (CTD 70) where sequences related to *Prasinoderma* (order Prasinococcales) were obtained (OTU Pras 2; see Figures 4.10). In both the northern gyre (CTD 25) and southern temperate (CTD 91) regions, sequences related to Pras 16S–VIII accounted for ~ 20 % of the clones (OTUs Pras 1; see Figures 4.10).

Other PPE classes which were contributors to the clone libraries at the 55% light level were Pelagophyceae, Trebouxiophyceae and Dictyochophyceae. Sequences phylogenetically affiliated within the class Pelagophyceae were observed in both Euk-A and Euk-B populations in the northern temperate region (CTD 12). Pelagophytes accounted for ~20% of clones in the Euk-B population in the southern gyre station (CTD 70) with sequences most closely related to *Pelagomonas calceolata* (see Figures 4.11). Clones related to Trebouxiophyceae accounted for ~40% of the Euk-A clone library in the equatorial upwelling region (CTD 56). Trebouxiophyceae were also significant contributors in the southern gyre region (CTD 83), accounting for 44% of the clones.

**Table 4.6. Numbers of clones relating to specific OTUs in either the Euk-A or Euk-B population plastid 16S rRNA clone libraries constructed from the 55% light level along AMT18.**

Region	Northern temperate		Northern gyre	Equatorial upwelling	Southern gyre			Southern temperate	
Station	12		25	56	70		83	91	
Depth	12m		19m	10m	18m		17m	9m	
PPE group	Euk-A	Euk-B	Euk-A	Euk-A	Euk-A	Euk-B	Euk-A	Euk-A	Euk-B
Prymn 1	-	10	-	-	-	-	-	31	50
Prymn 2	39	14	2	11	6	-	4	1	-
Prymn 3	1	7	-	-	4	-	25	-	-
Prymn 6	-	2	14	1	-	-	-	2	-
Prymn 7	-	-	-	-	-	-	3	-	-
Chryso 1	7	-	22	9	24	24	2	-	-
Treb 1	-	-	-	15	-	-	27	-	-
Pelago 1	6	2	1	-	-	7	-	-	-
Pras 1	-	-	11	1	-	-	-	9	-
Pras 2	-	1	-	-	12	6	-	-	-
Dict 1	8	1	7	-	-	-	-	-	-
Eust 1	-	-	-	1	-	-	3	-	-
Strept 1	-	-	-	-	-	-	-	2	-

In contrast to the 55% light level, where Prymnesiophyceae were the dominant class, Prasinophyceae and Chrysophyceae dominated at the DCM. In the northern temperate region (CTD 12) Prasinophyceae accounted for 66% and 45% of the Euk-A and B clone libraries, respectively (Figure 4.13). The majority of the prasinophyte sequences at this station were related to Pras 16S–VIII (OTU Pras 1) with a further proportion clustering in the order Prasinococcales (OTU Pras 2) (see Table 4.7 and Figure 4.10). A similar composition of the prasinophyte component was observed in the northern gyre station (CTD 25). In the equatorial upwelling region (CTD 56) the prasinophyte component, OTU Pras 2, was related to *Prasinoderma* (Prasinococcales). In the southern gyre (CTD 70) sequences clustered in the order Prasinococcales (OTU Pras 2) dominated the Euk-A clone library. In the second

southern gyre station (CTD 83), Prasinophyceae accounted for 20% of the clone library. The sequences were related to both Prasinococcales (OTU Pras 2) and Pras 16S–VIII (OTU Pras 1). In the southern temperate region (CTD 91) sequences related to Prasinococcales (OTU Pras 2) were observed in the Euk-A clone library.

In the equatorial upwelling region Chrysophyceae accounted for 70 and 78% of the clones in the Euk-A and B clone libraries, respectively (Figure 4.13). In the southern gyre station (CTD 83) Chrysophyceae accounted for 51% of the Euk-A clone library. Chrysophyceae were also significant contributors in the southern temperate region (CTD 91), accounting for 75% of the Euk-A clone library. At the other stations chrysophytes only made a minor contribution to the clone libraries. The majority of the chrysophyte sequences were loosely related to *Ochromonas distigma*.

At the DCM the class Prymnesiophyceae was only a minor component of the clone libraries except in the Euk-B population in the southern temperate region (CTD 91) where all the clones were related to *Chrysochromulina* B1 (Figure 4.13). Bacillariophyceae, which were not observed at the 55% light level, were present at the DCM in the northern gyre (CTD 25), equatorial region (CTD 56) and the southern gyre (CTD 83), but never exceeded more than 18% of the clones in any one library. Dictyochophytes were found only in the southern gyre (CTD 83) at the DCM.

In the northern gyre (CTD 25) 25% of the clones comprised sequences related to Trebouxiophyceae which were similar to those found in the surface at station 83. Pelagophyceae made a significant contribution (40%) to the Euk-B population in the northern temperate region (CTD 12) related to *Aureococcus anophagefferens* and *Pelagomonas calceolata*.

**Table 4.7. Numbers of clones relating to OTUs in either the Euk-A or Euk-B population plastid 16S rRNA clone libraries constructed from the DCM along AMT18.**

Region	Northern temperate		Northern gyre	Equatorial upwelling		Southern gyre		Southern temperate	
Station	12		25	56		70	83	91	
Depth	95m		120m	70m		150m	115m	41m	
PPE group	Euk-A	Euk-B	Euk-A	Euk-A	Euk-B	Euk-A	Euk-A	Euk-A	Euk-B
Prymn 1	-	-	-	-	-	-	-	4	50
Prymn 2	3	-	1	-	-	-	-	-	-
Prymn 3	2	-	-	-	-	-	-	-	-
Prymn 6	2	1	4	-	-	-	1	-	-
Chryso 1	6	3	4	150	124	11	20	25	-
Treb 1	-	-	11	-	-	-	-	-	-
Pelago 1	-	14	-	-	-	-	-	-	-
Pras 1	16	14	23	-	-	-	2	1	-
Pras 2	10	2	1	54	35	34	6	4	-
Bacillario 1	-	-	-	5	-	-	7	-	-
Bol 1	-	-	5	4	-	-	3	-	-
Strept 1	-	1	-	-	-	-	-	-	-

The Morisita Horn Similarity Index was used to assess the similarity in composition between the Euk-A and B populations. In the southern temperate region (CTD 91) the two populations had a high similarity at the 55% light level (0.907). At the DCM there was a low level of similarity (0.149), with the Euk-A fraction being dominated by chrysophytes while the Euk-B fraction was dominated by prymnesiophytes. In the northern temperate region the Euk-A and B populations had a high similarity in composition at the 55% light level (0.716). At the DCM a slightly lower level of similarity was observed (0.643). Other comparisons during AMT18 showed a high degree of similarity in the equatorial upwelling (CTD 56) and southern gyre (CTD 70) with values of 0.993 and 0.898 respectively.

The Morisita Horn Index was also used to compare the composition of the clone libraries between the two depths analysed. In the northern temperate region (CTD12) 8 OTUs were observed in the Euk-A population. Of these, only 3 OTUs were shared between the two depths. This gave a low similarity result of 0.188. For the Euk-B population only 3 out of 10 OTUs were shared between the two depths. A similarity value of 0.083 was obtained for the Euk-B size fraction. In the northern gyre (CTD 25) 9 OTUs were observed in the libraries but only 4 were shared between the two depths. A Morisita Horn value of 0.511 was achieved for the northern gyre region. In the equatorial upwelling region (CTD 56) there was little overlap in species composition with a Morisita Horn value of 0.387. In the southern gyre (CTD 70) only 2 out of 4 OTUs were shared between the two depths for the Euk-A population. This gave a Morisita Horn Similarity index value of 0.652. In the second southern gyre station (CTD83) a low level of similarity (0.05) was observed in the Euk-A population between the two depths. In the southern temperate region (CTD 91) the Euk-A population had a low similarity in composition (0.159) between the two depths. However the composition for the Euk-B population was identical being entirely composed of Prymnesiophyceae of the same OTU group (Prymn OTU 1).

The Morisita Horn similarity index was used to assess the similarity of samples along the transect, using the OTUs as a basis for investigation. Analysis of the Euk-A population at the 55% light level showed that the two temperate stations (CTD12 and 91), at the extremities of the cruise, had a low degree of similarity ( $C_{MH} = 0.029$ ). Although both these CTDs were dominated by prymnesiophytes CTD12 was dominated by *Chrysochromulina* clade B2 whilst *Chrysochromulina* clade B1 accounted for the majority of CTD91. CTD 12 and 91 had low similarities to the oligotrophic gyres (CTD 25, 70 and 83). A comparison between the temperate CTDs

(12 and 91) and the equatorial upwelling region (CTD 56) showed that there was a higher similarity than observed with the gyre CTDs. Comparisons between the North Atlantic gyre CTD 25 and the other CTDs showed that the highest similarity was observed with the southern gyre CTD 70 ( $C_{MH} = 0.655$ ). The Euk-A size fraction, CTD 83, at 17m had low similarities to any other Euk-A size fraction at the 55% light level (Table 4.8).

**Table 4.8. Morisita Horn Similarity Index between the CTDs at the 55% light level for the Euk – A population along AMT18.**

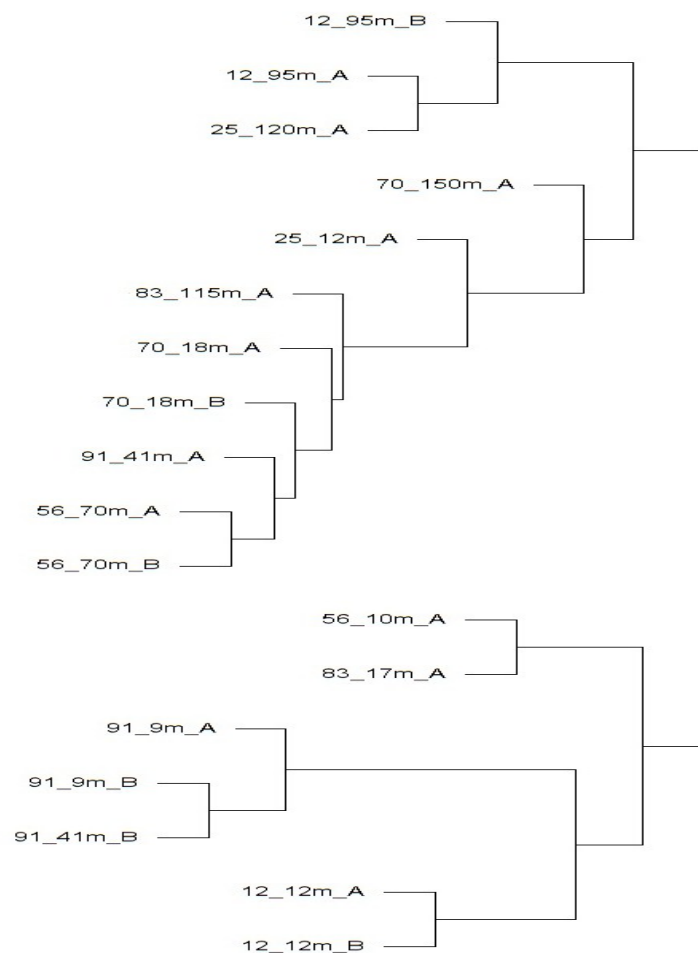
	12	25	56	70	83
25	0.237				
56	0.568	0.403			
70	0.355	0.655	0.486		
83	0.126	0.047	0.606	0.165	
91	0.29	0.128	0.031	0.006	0.003

Similarity in composition of Euk-A populations at the DCM along AMT18 was also analysed using the Morisita-Horn similarity index (Table 4.9). A different pattern was observed compared with that of the 55% light level. The northern temperate region had its highest similarity with the northern gyre ( $C_{MH} = 0.766$ ). The equatorial upwelling, CTD 56, had its highest similarity to the southern temperate CTD 91 ( $C_{MH} = 0.606$ ) (Table 4.9).

**Table 4.9. Morisita Horn Similarity Index between the CTDs at the DCM for the Euk-A population along AMT18.**

	12	25	56	70	83
25	0.766				
56	0.417	0.150			
70	0.514	0.076	0.610		
83	0.471	0.255	0.912	0.504	
91	0.370	0.176	0.969	0.447	0.884

A euclidean distance matrix of the stations was created using the different compositions of OTUs to show the similarity in species composition. Hierarchical cluster analysis was undertaken, using the distance matrix, in order to highlight the difference between the Euk-A and Euk B populations and CTDs. The dendrogram (Figure 4.14) highlighted the similarity between the two temperate regions, with the Euk-A and B fractions at the 55% light levels from both CTD12 and 91 clustering close to each other.



**Figure 4.14 Dendrogram depicting the composition of OTUs at various locations on AMT18 (CTD number) in the Euk-A and Euk-B PPE populations created using a Euclidean distance matrix.**



#### **4.3.4 Comparison of PPE community structure as assessed using environmental DNA extracted from bulk community filters or from flow cytometrically sorted cells.**

Using marine-algal plastid biased 16S rRNA gene primers and the datasets obtained in sections 4.3.2 and 4.3.3 above, a comparison of PPE diversity between filtered samples and flow cytometry sorted cells was undertaken. Clearly evident was that sequences related to the class Prymnesiophyceae were the dominant components of the PPE community using both approaches. Such sequences were related to *Chrysochromulina* clades B1 and B2, the Prym 16S–II clade and the order Coccolithales. However, sequences relating to Phaeocystales were only obtained in the filtered samples.

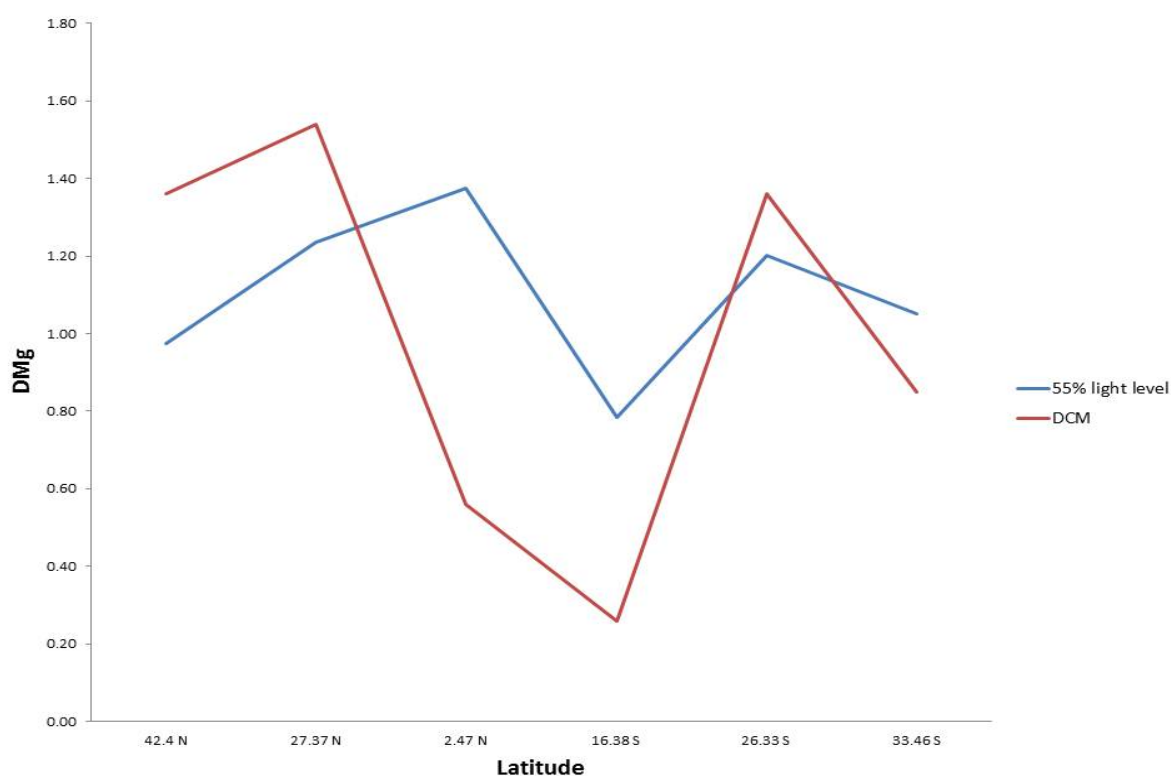
Members of the Chrysophyceae were present in both the filtered and sorted samples. In the filtered samples the abundance of chrysophytes was seen to diminish with depth. However in the sorted samples, chrysophyte sequences related to *Ochromonas distigma* were significant components in the DCM samples especially at CTD 56.

Prasinophyceae were also present in the libraries created using both methods. In the sorted cells, sequences related to Prasinococcales and Pras 16S–VIII were present. Sequences relating to Pras 16S–VIII were also present in the filtered samples as well as sequences relating to Prasinophyceae clade VIIA and Pryamimonadales.

Dictyochophyceae and Pelagophyceae were observed in samples obtained using both methods. However, Trebouxiophyceae and Bacillariophyceae were only observed in sorted samples. Conversely sequences affiliated with Bolidophyceae, Rhodophyta and rappemonads were only observed in the filtered samples.

### 4.3.5 Species richness of flow cytometrically sorted clone libraries along the AMT18 latitudinal gradient

Species diversity was compared along the transect using the Margalef diversity index. At the 55% light level a peak in diversity was observed near the equator, with diversity decreasing with increasing latitude. A dip in diversity was also observed at 16°S, where the clone library was dominated by sequences corresponding to *Prasinococcales* (Figure 4.15).



**Figure 4.15** Margalef's diversity indices at the 55% light level and DCM along AMT18.

At the DCM there was a difference in the trend of diversity along the transect compared with the 55% light level. The peak in diversity was observed at the northern gyre station (CTD 25) at 27°N. The lowest diversity was observed at 16.38°S.

In surface waters (i.e. at the 55% light level) Pearson correlation coefficients showed that there was a negative relationship between these diversity indices and latitude (-0.385), although this relationship was not statistically significant ( $p > 0.1$ ) (Table 4.10). There were also negative relationships between diversity, salinity and chlorophyll *a*, which were also not statistically significant. Non significant positive correlations were observed between diversity, temperature, phosphate, and silicate concentrations.

**Table 4.10. Pearson correlation coefficients for environmental variables with Margalef's diversity values at the 55% light level and DCM. Significant results ( $p < 0.05$ ) are highlighted in bold.**

	55%	DCM
Latitude	-0.385	0.647
Temperature	0.446	-0.521
Salinity	-0.491	0.083
Chlorophyll	-0.106	-0.474
Phosphate	0.456	-0.426
Silicate	0.553	-0.117
Nitrite	NA	<b>-0.852</b>
Nitrate	NA	0.225

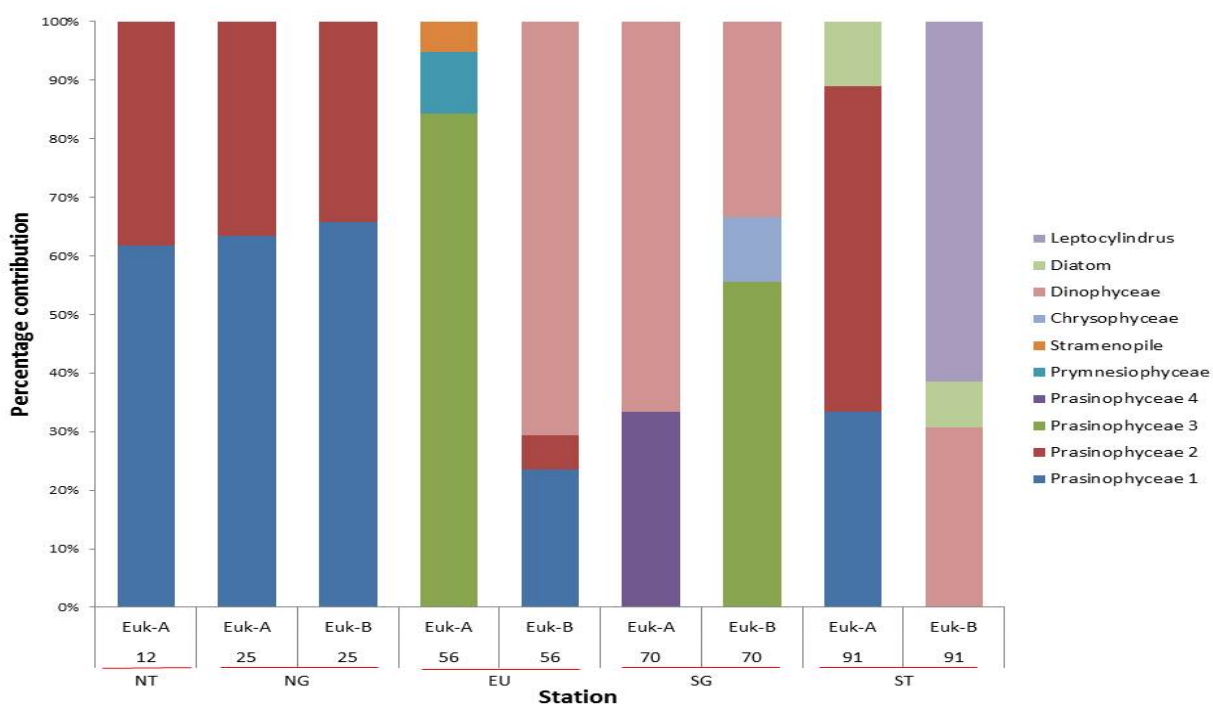
At the DCM, Pearson correlation coefficients showed that there was a positive correlation between diversity and latitude, although this was not statistically significant (Table 4.10). However, there was a significant negative correlation between diversity and nitrite concentration ( $p < 0.05$ ). A negative correlation was also observed between diversity and temperature, although this relationship was not significant.

#### **4.3.6 The composition of picoeularyotes using 18S rRNA gene clone libraries from flow cytometry sorted cells**

18S rRNA gene libraries were created from flow cytometry sorted cells at the 55% light level (see section 4.2.7) at 5 stations along AMT18 corresponding to CTDs 12, 25, 56, 70 and 90.

Prasinophyceae were the dominant photosynthetic component throughout AMT18 in these 18S rRNA gene clone libraries (Figure 4.16). In the northern temperate region (CTD 12) the library was comprised of 2 OTUs relating to photosynthetic organisms. The corresponding sequences were phylogenetically affiliated to Prasinophyceae clade VIIA (OTU 2) and clade VIIB (OTU 1) (see Figure 4.17). In the northern gyre region (CTD 25) both the Euk-A and -B populations comprised sequences again related to Prasinophyceae clade VIIA and clade VIIB. In the equatorial upwelling region (CTD 56) sequences affiliated with Mamiellophyceae accounted for 88% of the Euk-A clone library (OTU 3).

The dominance of prasinophytes declined in the Euk-B library, only accounting for 30% of the library. In the southern gyre (CTD 70) Prasinophyceae contributed only 7% of the Euk-A community, though this was attributable to a single OTU (OTU 4) the corresponding sequencing of which showed phylogenetic affiliation to Pyramimonadales (Table 4.11). In the Euk-B population prasinophytes (corresponding to OTU 3) made a greater contribution, accounting for 35% of the clone library related to Mamiellophyceae. In the southern temperate region (CTD 91) the Euk-A population was again dominated by Prasinophyceae clade VIIA and clade VIIB. However, in the Euk-B population no Prasinophyceae sequences were observed.

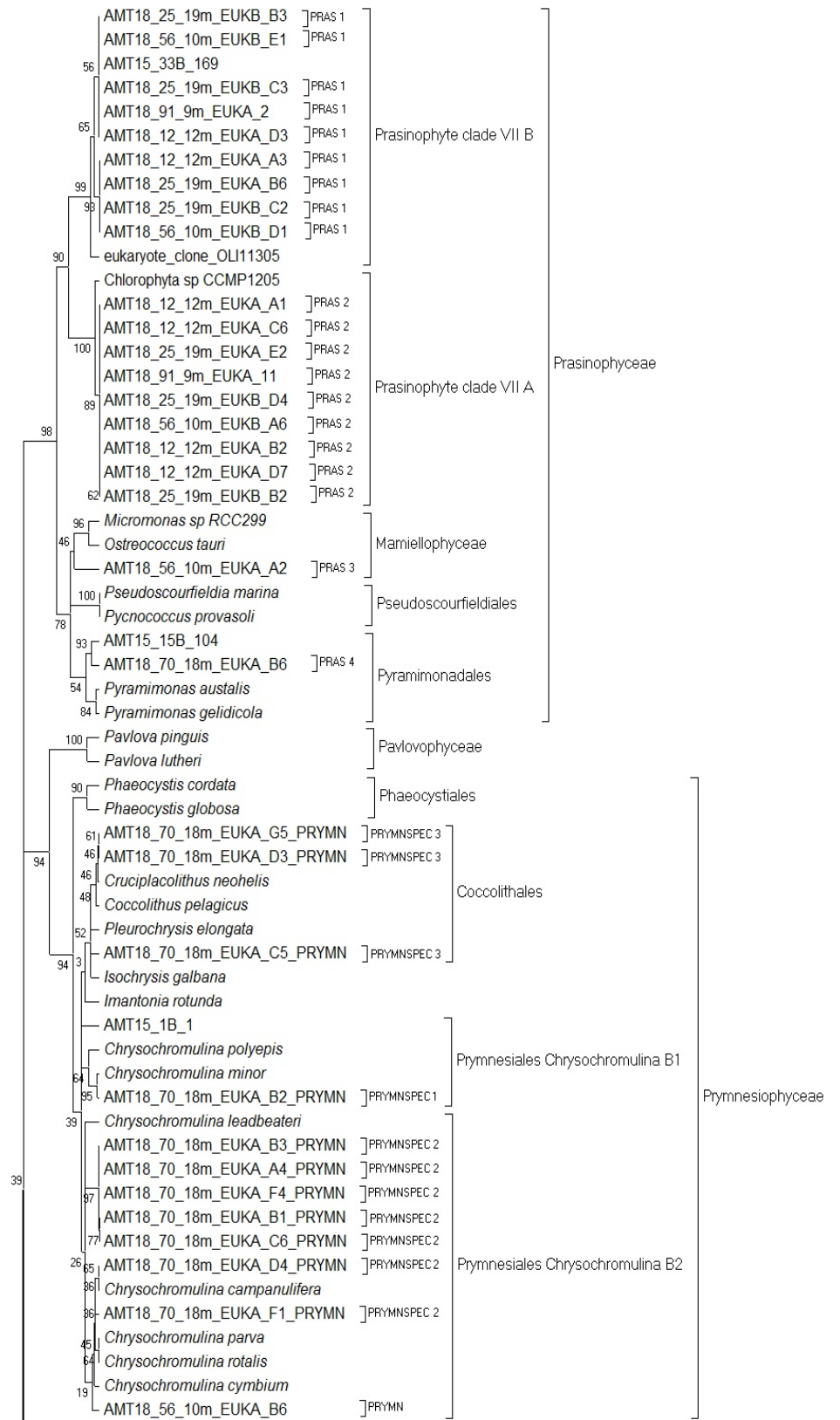


**Figure 4.16.** PPE composition of AMT18 stations from OTUs derived from nuclear 18S rRNA gene libraries.

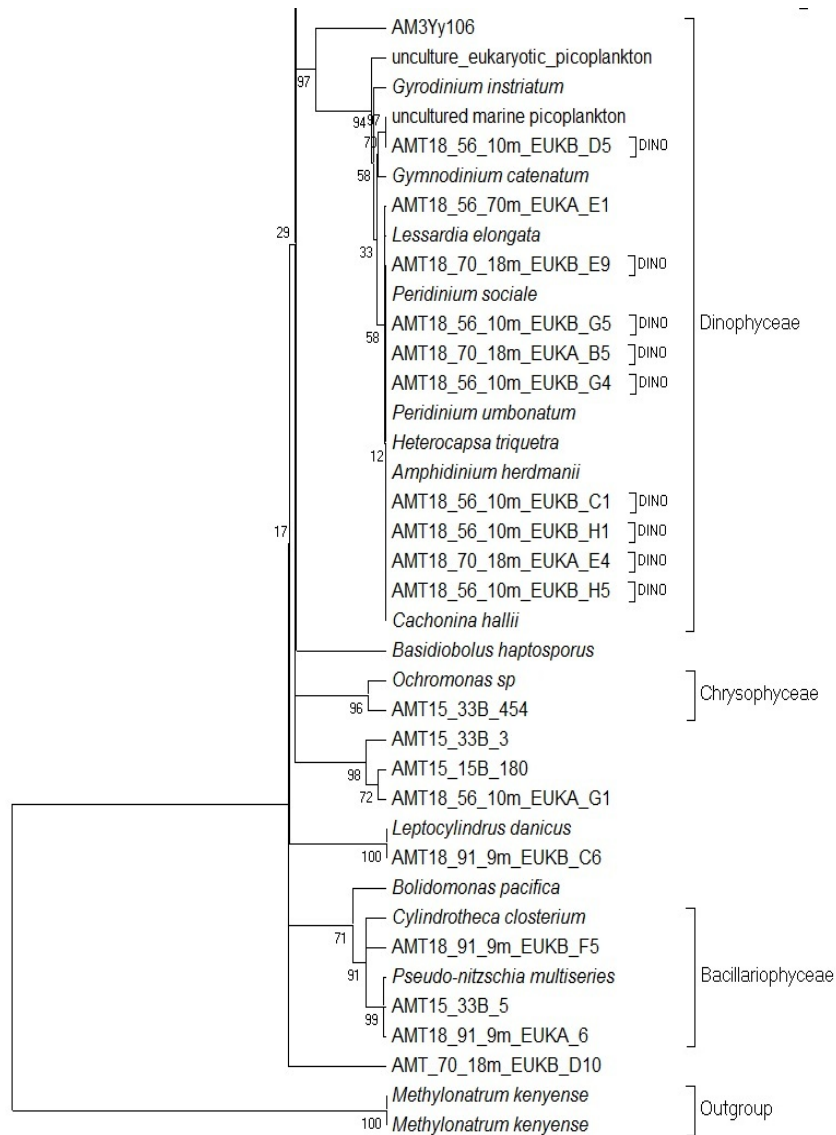
**Table 4.11.** Numbers of clones relating to OTUs in either the Euk-A or Euk-B population 18S rRNA gene clone libraries constructed from the 55% light level along AMT18.

Region	Northern Temp	Northern Gyre		Equatorial Upwelling		Southern Gyre		Southern Temp	
Depth	12	25	25	56	56	70	70	91	91
PPE group	Euk-A	Euk-A	Euk-B	Euk-A	Euk-B	Euk-A	Euk-B	Euk-A	Euk-B
Prasinophyceae 1	34	26	23	-	4	-	-	6	-
Prasinophyceae 2	21	15	12	-	1	-	-	10	-
Prasinophyceae 3	-	-	-	16	-	-	5	-	-
Prasinophyceae 4	-	-	-	-	-	1	-	-	-
Prymnesiophyceae	-	-	-	2	-	-	-	-	-
Stramenopile	-	-	-	1	-	-	-	-	-
Chrysophyceae	-	-	-	-	-	-	1	-	-
Dinophyceae	-	-	-	-	12	2	3	-	4
Bacillariophyceae	-	-	-	-	-	-	-	2	1
Leptocylindrus	-	-	-	-	-	-	-	-	8
Fungi	16	31	4	4	3	3	6	6	29
Syndiniales	-	-	2	2	0	11	8	4	16

The other algal class which significantly contributed to the 18S rRNA gene clone libraries was Dinophyceae. Dinophyceae were the dominant class in the Euk-B population in the equatorial upwelling region (CTD 56). This class was also observed in the southern gyre station (CTD 70) in both Euk-A and Euk-B populations and accounted for ~30% of the Euk-A clone library in the southern temperate region (CTD 91) with sequences related to *Peridinium* at both stations (see Figure 4.17). Minor contributions were made by Prymnesiophyceae in the Euk-A clone library in the equatorial upwelling (CTD 56) and the Euk-B population in CTD 70. The prymnesiophyte sequences were related to *Chrysochromulina* clade B2. Bacillariophyceae were only present in the southern temperate region (CTD 91) but accounted for almost 70% of the Euk- B population at this location.



**Figure 4.17. Phylogenetic tree of nuclear 18S rRNA gene sequences obtained in clone libraries constructed along AMT18.**



**Figure 4.17 cont.**

Non photosynthetic lineages were also obtained in the clone libraries. Sequences attributable to Syndiniales and fungi were observed throughout the transect. However, given the non-photosynthetic nature of these groups these Syndiniales and fungal sequences were removed from further analysis.



#### 4.3.7 A focused investigation of the diversity of Prymnesiophyceae at a single location in the southern gyre along AMT18 using class-specific primers on flow cytometry sorted Euk-A cells

The specific composition of Prymnesiophyceae in the southern gyre in the Euk-B PPE population (CTD70, depth 18m) was assessed using the 18S rRNA gene and a nested PCR approach (see section 4.2.7). Thus, general 18S rRNA gene primers were used in a first round of amplification, followed by a second round using Prymnesiophyceae specific primers. Of the 45 clones which were analysed, 3 OTUs were identified relating to prymnesiophytes (Table 4.12). The coverage values of this library was 0.95, suggesting that a large proportion of the OTUs in this sample had been identified. The most dominant OTU present (OTU 18S Prymn 2), accounting for 40% of the library, clustered within *Chrysochromulina* clade B2 clade (Figure 4.17). In total, sequences related to *Chrysochromulina* clade B2 accounted for 75% of the library. The remainder of the library was made up of 6 clones (13%) related to *Chrysochromulina* clade B1 (OTU 18S Prymn 1) and a total of 5 clones (OTU 18S Prymn 1) clustering in the Coccolithales.

**Table 4.12. The number of clones and phylogenetic affiliation of OTUs obtained from the prymnesiophyte specific 18S rRNA gene clone library.**

OTU	Clone #	Phylogeny
18S Prymn 1	34	<i>Chrysochromulina</i> B1
18S Prymn 2	6	<i>Chrysochromulina</i> B2
18S Prymn 3	5	Coccolithales

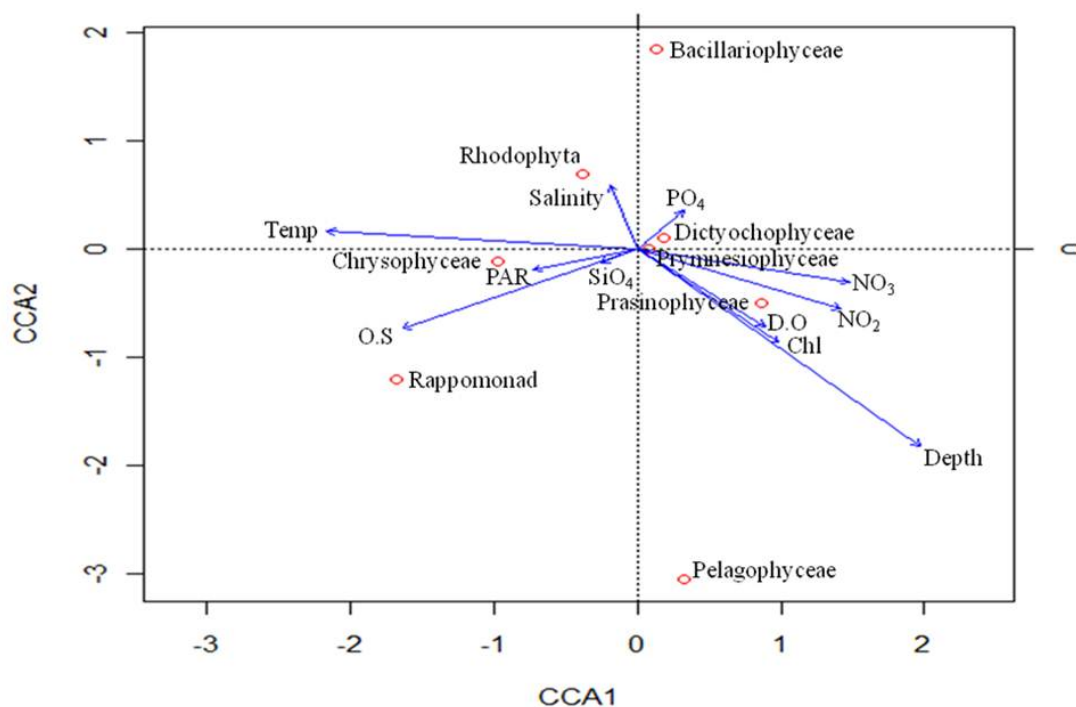
#### **4.3.8 Comparison of PPE diversity along AMT18 using the various plastid and nuclear targeted PCR primer sets**

Using the marine algal plastid biased 16S rRNA primers, Prymnesiophyceae and Chrysophyceae were the dominant classes in the AMT18 plastid clone libraries. In contrast, the nuclear 18S rRNA clone libraries showed a dominance of Prasinophyceae with members of the Dinophyceae also present. In these latter libraries the Prasinophyceae component comprised OTUs related to Prasinophyceae clade VIIA and clade VIIB whereas neither of these clades were observed in the plastid targeted libraries. Furthermore, in the nuclear 18S rRNA gene clone libraries Prymnesiophyceae clones were only obtained in small numbers, whilst in the corresponding plastid libraries they were the dominant class. However, when using Prymnesiophyceae-specific primers targeting the 18S rRNA gene, a similar prymnesiophyte diversity was obtained compared to that obtained with the plastid primers, with sequences attributable to *Chrysochromulina* clade B2 being the dominant clade in both cases.

#### **4.3.9 Canonical correspondence analysis (CCA) of PPE community structure determined from bulk community DNA samples and measured environmental variables along AMT18.**

CCA analysis was used to assess the association of measured environmental variables with the PPE class-specific distribution patterns determined from clone libraries constructed from bulk DNA samples (Figure 4.18). Distribution patterns of Prasinophyceae were strongly associated with dissolved oxygen, chlorophyll *a* concentration and depth. In contrast, Chrysophyceae distribution patterns were associated with PAR irradiance and oxygen saturation, whilst none of the measured

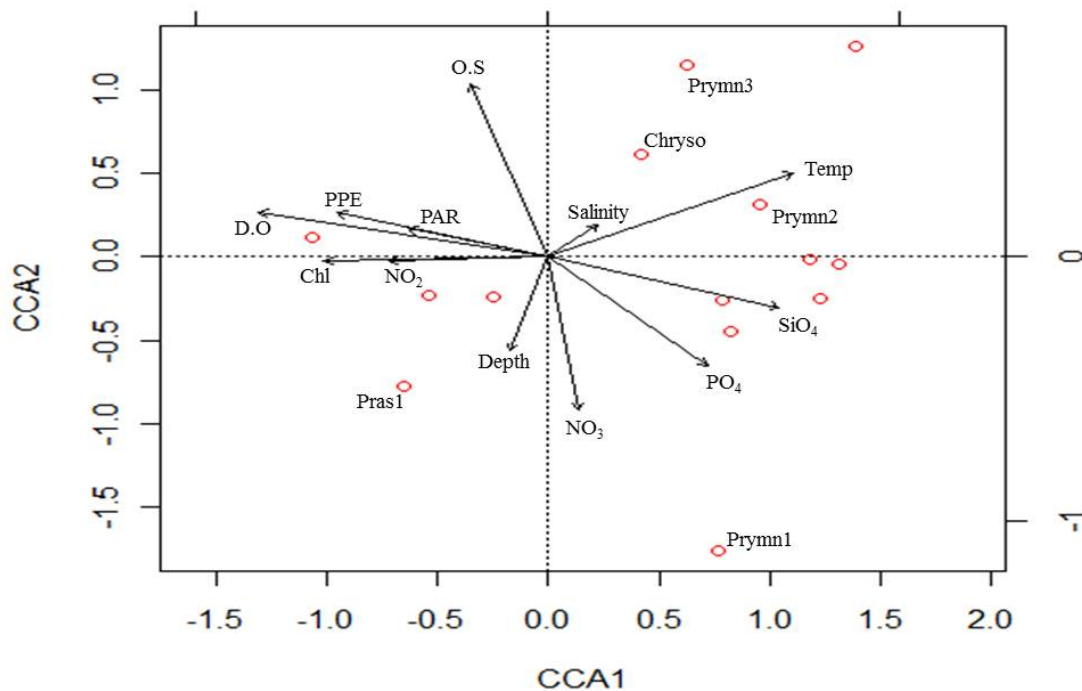
variables were able to explain the distribution of prymnesiophytes. Forward selection was used to identify the factors explaining a significant amount of the variation in PPE class distribution patterns, which were temperature and depth. However, these factors only explained 18.9% of the variation.



**Figure 4.18. CCA plot for PPE class distributions along AMT18 in relation to measured environmental variables, using data from clone libraries constructed from bulk community DNA. Classes are depicted with red circles. Arrows pointing in roughly the same direction indicate a high positive correlation, arrows crossing at right angles indicate a near-zero correlation, and arrows pointing in the opposite direction have a high negative correlation. O.S = Oxygen saturation; D.O = Dissolved oxygen; Chl = chlorophyll**

CCA analysis was also used to assess the association of measured environmental variables with the distribution patterns of specific OTUs determined from plastid clone libraries constructed from bulk DNA samples (Figure 4.19). This showed that

*Chrysochromulina* clade B2 (OTU Prymn 2) had a positive association with temperature. Coccolithales (OTU Prymn 3) had a positive correlation with salinity and temperature. The Chrysophyceae OTU had a negative association with depth and a positive association with salinity and temperature. Prasinophyceae clade VIII (OTU Pras 1) had a positive correlation with depth.



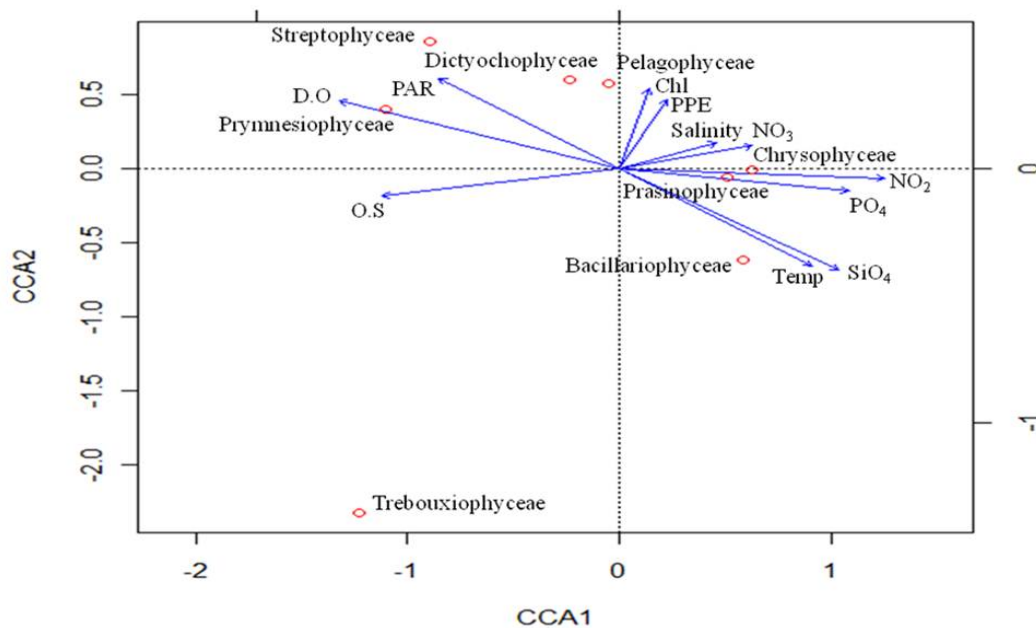
**Figure 4.19. CCA plot for specific PPE OTU distributions along AMT18 in relation to measured environmental variables, using data from clone libraries constructed from bulk community DNA. OTUs are depicted as red circles and dominant OTUs have been labelled. O.S = Oxygen saturation; D.O = Dissolved oxygen; Chl = chlorophyll**

#### **4.3.10 CCA of PPE community structure determined from flow cytometry sorted cells and measured environmental variables along AMT18.**

CCA analysis was also used to assess the association of measured environmental variables with the PPE class-specific distribution patterns determined from clone

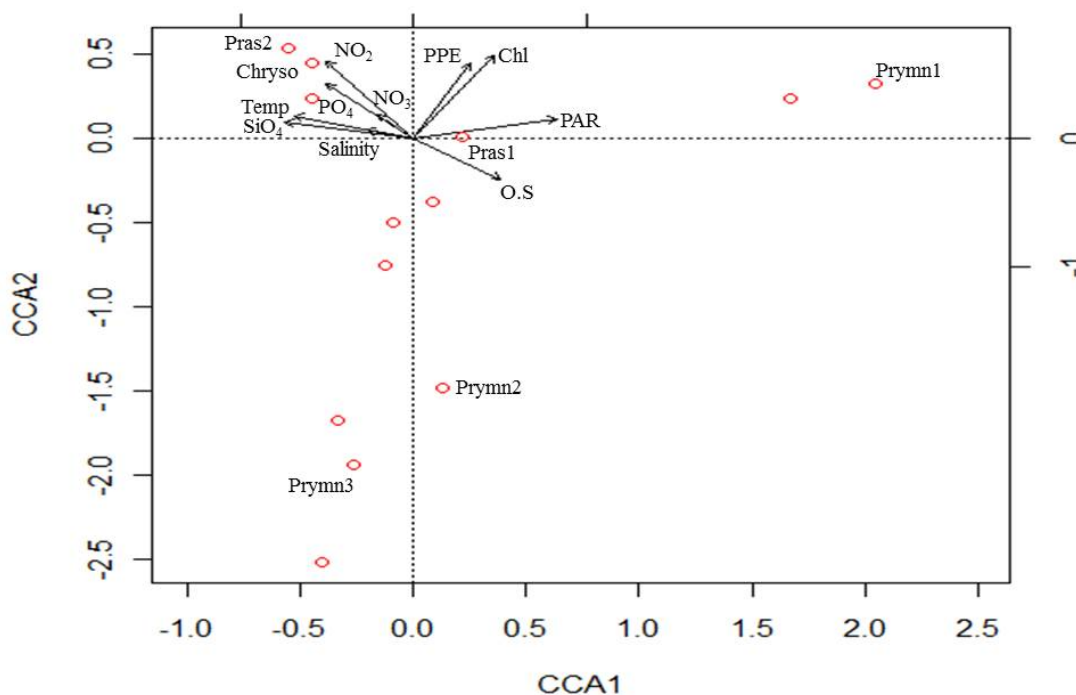
libraries constructed from flow sorted cells (Figure 4.20). This showed that Prasinophyceae and Chrysophyceae were associated with salinity and nitrate concentrations. Prymnesiophyceae had minor associations with dissolved oxygen and PAR irradiance, whilst the other important PPE class observed along AMT18, Chrysophyceae, had minor associations with concentrations of nitrite, phosphate and silicate, as well as temperature (Figure 4.20). Forward selection was used to select the environmental variables that explained a significant part of the variation in PPE distribution patterns ( $p < 0.05$ ). The factors which explained a significant proportion of the variance were nitrate, oxygen saturation, dissolved oxygen and light attenuation. Variation Partitioning Analysis (VPA) was performed, which indicated that only 42% of the variation in PPE diversity was explained by the measured variables.

Further CCA analysis was then performed to assess the association of measured environmental variables with the distribution patterns of specific OTUs determined from clone libraries constructed from flow sorted cells (Figure 4.21). The distribution of *Chrysochromulina* clade B2, the dominant prymnesiophyte clade along the transect, was poorly defined by any of the measured variables. *Chrysochromulina* clade B1 was associated with PAR irradiance as were members of Prasinophyceae clade VIII (Figure 4.21). Prasinophyte sequences related to Prasinococcales (OTU Pras 2) which were significant components at the DCM at CTD 56 and CTD 70, were associated with the nutrients phosphate, nitrate, nitrite and silicate, as well as temperature. The chrysophyte OTU which was particularly abundant at stations CTD25, 56 and 70, was also associated with the nutrients nitrate, nitrite, silicate and phosphate, as well as temperature. Trebouxiophyte OTUs were associated with dissolved oxygen and oxygen saturation.



**Figure 4.20. CCA plot for PPE class distributions along AMT18 in relation to measured environmental variables, using data from clone libraries constructed from flow sorted cells. Classes are depicted with red circles. O.S = Oxygen saturation; D.O = Dissolved oxygen; Chl = chlorophyll**

Using forward selection, the environmental variables that explained a significant part of the variation in PPE distribution, were selected ( $p < 0.05$ ). VPA was performed using these variables which indicated that ~59% of the variation in PPE diversity was explained by these variables. The nutrients, nitrite and silicate accounted for ~11.5% of this variation whilst temperature, PAR irradiance, PPE abundance, chlorophyll *a* and dissolved oxygen accounted for ~35% with the remaining ~3% explained by a combination of the variables. A higher proportion of the variation was explained when using OTUs, than when the analysis was undertaken at the class level.



**Figure 4.21.** CCA plot for specific PPE OTU distributions along AMT18 in relation to measured environmental variables, using data from clone libraries constructed from flow sorted cells. Classes are depicted with red circles. O.S = oxygen saturation.

## **4.4 Discussion**

### **4.4.1 PPE abundance along AMT18**

In the permanently stratified tropical ocean, the surface is characterised by a low biomass of phytoplankton, whilst a disproportionate amount of phytoplankton absorbance and production occurs at a sub surface chlorophyll *a* maximum (DCM). The upper layers of stratified oceanic waters are characterised by nutrient depletion, which renders primary production dependent on nutrient flux across the thermocline. However, whilst the bottom of the mixed layer is an appropriate environment for nutrition, the lower light levels exert a constraint on the organisms inhabiting this

environment and there is a trade off between nutrition and the shaded nature of the environment (Agusti and Duarte, 1998). The distribution of PPEs along AMT18, which mirrors that of chlorophyll *a*, is similar to that seen on previous AMT transects (Tarran *et al.*, 2006), with abundance counts comparable to those found during AMT13 and AMT14. The abundance of the Euk-A population was higher than the Euk-B population throughout AMT18. The numerical dominance of the Euk-A population in these open ocean waters is in agreement with remote sensing models, which suggest that organisms <2 µm are dominant, whilst larger size fractions predominate in more coastal mesotrophic regions (Hirata *et al.*, 2008). The relationship between cell size and growth rate has generally been shown to be inverse, due to thinner diffusion boundaries and greater surface area-volume ratios. Due to the necessary occupation of a larger fraction of the biomass by nonscalable components, such as genome and various membranes, there are constraints on the maximum growth rate such that below a certain size growth rate will become positively correlated with cell size. Indeed, Bec *et al.*, (2008) demonstrated that for eukaryotic algae the maximum growth rate was achieved in the 2-3 µm size fraction.

#### **4.4.2 Diversity of PPEs using marine algal plastid biased 16S rRNA PCR primers**

Prokaryotic 16S rRNA gene sequences have been extensively used to study microbial diversity in both cultures and natural environments (Fuller *et al.*, 2003, Rocap *et al.*, 2002). However, PPEs are poorly reflected in clone libraries using universal bacterial primers, generally due to the low numerical abundance of the PPEs, especially in the oligotrophic open ocean, compared to heterotrophic bacteria and picocyanobacteria (Moreira and Lopez-Garcia, 2002). The poor representation



of PPEs in 16S rRNA libraries led Fuller *et al.*, (2006a) to develop a marine algal plastid biased primer set in order to investigate PPE diversity in the environment. Use of such primers in a study in the Arabian Sea (Fuller *et al.*, 2006a), demonstrated their efficacy in detecting most PPE classes, including the classes Chrysophyceae, Prymnesiophyceae, Pelagophyceae, Dictyochophyceae and Prasinophyceae. However, the class Mamiellophyceae appears to be more poorly represented using this primer set (McDonald *et al.*, 2007).

The dominance of sequences related to Prymnesiophyceae in the plastid clone libraries is in agreement with previous pigment studies. The dominant pigment along AMT 2-5 was 19' hexanoyloxyfucoxanthin, a pigment distinctive of Haptophyta (Gibb *et al.*, 2000). Using FISH analysis, coupled with radiotracer uptake and flow cytometric sorting pico sized prymnesiophytes have recently been shown to be major players in CO<sub>2</sub> fixation in the subtropical and tropical northeast Atlantic (Jardillier *et al.*, 2010). In the sorted samples, CCA analysis showed that Prymnesiophyceae had a minor negative association with temperature due to prymnesiophyte sequences dominating the libraries in the northern and southern temperate regions. In the filtered samples none of the measured factors explained the distribution of prymnesiophytes along the transect. CCA analysis showed that the OTU related to *Chrysochromulina* clade B2 (OTU Prymn 2) and Coccolithales (OTU Prymn 3) were not explained by any variables measured. However, *Chrysochromulina* clade B1 (OTU Prymn 1) was associated with PAR irradiance. This is in accordance with these sequences being observed at the 55% light level.

A complimentary distribution pattern was observed between Chrysophyceae and Prymnesiophyceae during AMT18. Prymnesiophyceae dominated the clone libraries

at the extremities of the cruise, whilst Chrysophyceae were more dominant in the equatorial and southern gyre regions. The pattern of distribution of these two classes is similar to that found during AMT15 (see Kirkham *et al.*, 2011).

In the current study, in the filtered samples, Chrysophyceae were an important component of the samples and there was also a decline in abundance of clones corresponding to this class with depth. This observation is in agreement with previous studies. During AMT15 Kirkham *et al.*, (2011) also observed a decline of chrysophytes with depth. However, in the sorted samples, Chrysophyceae sequences were just as abundant at the DCM as they were higher up in the water column. CCA showed that chrysophytes were associated with light levels and oxygen saturation in the filtered samples which is a similar trend to that observed during a previous AMT (Kirkham *et al.*, 2011). However, for the sorted cells, CCA showed Chrysophyceae no longer had an association with light but had an association with the nutrients nitrite, phosphate and silicate, as well as temperature. The order Parmales, was long thought the only documented group within the Chrysophyceae with a marine distribution, comprising solitary cells 2-5  $\mu\text{m}$  in diameter, with a single chloroplast and a silicified cell wall (Bravo-Sierra and Hernandez-Becerril, 2003). This requirement for silicate in the cell wall potentially could have explained the association with silicate in the CCA for sorted cells. However, very recently a species of Parmales was isolated into culture and phylogenetic analysis showed these parmalean algae are in fact within the bolidophycean clade of autotrophic naked flagellates (Ichinomiya *et al.*, 2011) and not related to Chrysophyceae at all. Characterisation of the organisms relating to the plethora of Chrysophyceae-related sequences found in the plastid-derived clone libraries here and elsewhere (e.g. see Fuller *et al.*, 2006a; Lepère *et al.*, 2009) thus remains elusive.

The plastid rRNA gene libraries indicated that Prasinophyceae were also significant contributors to the PPE community along AMT18, especially in the flow sorted samples (Figures 4.12 and 4.13). The smaller contribution of prasinophytes in filtered samples could be due to PCR bias against some prasinophyte groups (e.g. Prasinophyceae clade II, now re-classified as Mamiellophyceae (Marin and Melkonian 2010), which have been documented for the plastid primers used here (McDonald *et al.*, 2007). No sequences related to Mamiellophyceae were observed during AMT18. However, Mamiellophyceae are mostly known as important components of the PPE community in coastal regions (see e.g. Not *et al.*, 2004, 2008). Hence, it is not so surprising that they are poorly represented in these open ocean clone libraries, even given the apparent bias of the PCR primer set used. Phylogenetic analysis of prasinophyte sequences revealed that sequences related to Prasinococcales were significant contributors to the clone libraries, especially in the southern gyre (CTD 70) (Figure 4.13). CCA analysis indicated an association with higher nutrients and temperature for Prasinococcales OTUs (OTU Pras 2). The association with higher nutrient status is in agreement with previous studies, which found this order restricted to mesotrophic regions of the Pacific (Lepère *et al.*, 2009) and with isolation studies (Le Gall *et al.*, 2008). The other group which was revealed in the sorted samples was related to Pras 16S-VIII (Lepère *et al.*, 2009), a novel lineage which currently contains no cultured representatives but are well represented in libraries, across a range of trophic regimes, constructed from the southeast Pacific Ocean (Lepère *et al.*, 2009).

CCA showed that the measured variables could only explain between 21-59 % of the variation in the clone library data. The remaining variation could be explained for example by biotic factors, such as viral lysis and predator abundance/grazing rates,

or other chemical factors, such as iron concentration, which were not measured in this study. Viral lysis of marine eukaryotic phytoplankton has been shown to be a controlling factor. Viruses have been shown to be capable of infecting the major planktonic groups including diatoms, prasinophytes, chrysophytes and prymnesiophytes (Fuhrman, 1999). High abundance of mixotrophs in the marine environment as observed by Zubkov and Tarran (2008) in the North Atlantic may suggest that the availability of bacterioplankton could influence the distribution of picoeukaryotes. Top down factors such as zooplankton grazing have been shown to have an impact on the composition of small eukaryotes in both freshwater and marine environments (Lepère *et al.*, 2006 and Baudoux *et al.*, 2008). In the nuclear 18S rDNA clone libraries sequences related to the parasitic order Syndiniales were observed in the flow cytometry sorted libraries (their presence is perhaps a result of them infecting PPE cells directly, or more likely due to them being co-sorted with PPEs). These organisms have been shown to have a wide distribution in the marine environment (Guillou *et al.*, 2008). Syndiniales have been shown to control blooms of toxic marine dinoflagellates and they may thus have an impact on the composition of marine phytoplankton (Chambouvet *et al.*, 2008). The low amount of variation explained by the factors measured suggests that the distribution of PPEs may be more influenced by other factors such as grazing/infection rates and prey availability.

#### **4.4.3 Diversity of PPEs along AMT18 as evidenced using nuclear 18S rRNA gene primers.**

In the current study, sequences from two OTUs, phylogenetically affiliated with Prasinophyceae clades VIIA and VIIB, made a significant contribution to the clone libraries in the northern temperate and northern gyre regions. Sequences

corresponding to these two clades were also observed in the equatorial upwelling region. Guillou *et al.*, (2004) investigated the diversity of picoplanktonic prasinophytes in the marine environment; they found that Prasinophyceae clade VII formed three distinct lineages. Clade VIIC contained the cultured representative *Picocystis salinarum*, which is found in hypersaline environments but has not been observed in marine environments (Guillou *et al.*, 2008). The other two lineages have been observed in marine systems. Clade VIIA contains environmental sequences from a diverse range of environments, including the equatorial Pacific and coastal regions and is also represented in culture. However, thus far, Prasinophyceae Clade VIIB is only composed of environmental sequences from the equatorial Pacific (Guillou *et al.*, 2004; Shi *et al.*, 2009). Here, sequences related to Prasinophyceae clade VIIB had a wide distribution along AMT18. The presence of these sequences in the mesotrophic temperate regions and oligotrophic gyre regions is similar to the distribution observed by Guillou *et al.*, (2004).

In the equatorial upwelling, sequences phylogenetically affiliated to *Pryamimonas* were dominant in the library, although they were only found in one other CTD. This genus has previously been sporadically isolated from coastal environments (Vaulot *et al.*, 2008). In coastal waters Romari and Vaulot (2004) found that, among the Prasinophyceae, most of the sequences corresponded to three genera, *Micromonas*, *Ostreococcus* and *Bathycoccus*, all members of the class Mamiellophyceae. This is in agreement with previous studies which have found that sequences related to this class were only found in upwelling coastal regions (Shi *et al.*, 2011).

The other significant class observed in the clone libraries were the Dinophyceae. Despite the fact that no species with a minimum size <3µm have yet been described,

this group is second in abundance among sequences which have been filtered through 3  $\mu\text{m}$ , although a significant proportion of these may originate from heterotrophic species (Vaulot *et al.*, 2008). However, their relative abundance in the euphotic zone suggests that some may be autotrophic.

The use of flow cytometry has enabled a more targeted approach for characterising the diversity of natural PPE populations, as cells can be discriminated into specific subpopulations. However, until recently, such as Marie *et al.*, (2010), studies have not utilised flow cytometry to assess the diversity of picoeukaryotes. In the present study sequences related to the non photosynthetic lineage, Syndiniales, were observed in the clone libraries in low numbers. The use of flow cytometry has reduced the bias towards obtaining heterotrophic sequences, which has previously been observed in 18S rRNA clone libraries. In recent studies on filtered samples, undertaken in waters off Roscoff (Brittany, France), only a few sequences were obtained from photosynthetic lineages such as Prasinophyceae, whilst the libraries were dominated by heterotrophic organisms such as marine stramenopiles (MAST, see Massana *et al.*, 2004) and Syndiniales (alveolate groups 1 and 2, see Guillou *et al.*, 2008). Vaulot *et al.*, (2002) recovered only two autotrophic organisms amongst 145 sequences obtained at a coastal site in the English Channel. This suggests that the diversity of autotrophs may be lower than that of heterotrophs. This could be due to the possibility of fewer niches being available for autotrophs than heterotrophs, especially in coastal waters where organic matter is more abundant and diverse (Vaulot *et al.*, 2002). Heterotrophic organisms are involved in the degradation of a variety of organic molecules, thus allowing for various niches based on the specialisation of degradation to particular organic molecules (Vaulot *et al.*, 2002). The dominance of non photosynthetic lineages in this area is surprising as FISH

results suggest that 85% of the eukaryotic cells belong to Chlorophyta (Not *et al.*, 2004).

All of the 18S rDNA clone libraries created during this study had sequences relating to fungi present. This is a potential problem for the targeting of picoeukaryotes when using flow cytometry, as the amount of material recovered is quite low and is prone to contamination, in particular from fungi (Marie *et al.*, 2010). In this study, between 25,000 and 100,000 cells were sorted due to limitations in the number of cells concentrated especially in the gyre regions. Marie *et al.*, (2010) found that sorted populations of 100,000 cells for picoeukaryotes gave a reliable clone library construction, with minimal contamination without relying on nested PCR. The comparatively low number of cells sorted in this study may explain the presence of fungal sequences in the clone libraries. Conversely they may be *bone fide* sequences worthy of further investigation.

In the current study chimeras accounted for between 5-52 % of the clones in the libraries made, with the highest numbers being observed in clones derived from the Euk-B population. The formation of chimeras in 18S rRNA gene clone libraries has recently received some attention (Berney *et al.*, 2004). Marie *et al.*, (2010) also found that chimeras were abundant in their data and proposed that chimera formation containing closely related organisms could be higher in low diversity populations and that primer sets may also have an influence.

#### **4.4.4 Comparison of PPE community structure along AMT18 using plastid and nuclear rRNA clone libraries on flow cytometrically sorted cells.**

Differences in the composition of the clone libraries constructed using plastid biased 16S rRNA and nuclear 18S rRNA primers were observed throughout the transect.

The dominant classes in the marine algal plastid biased 16S rRNA gene libraries were Prymnesiophyceae and Chrysophyceae, with Prasinophyceae also forming a significant proportion of the clone libraries. In the nuclear 18S rRNA gene clone libraries, Prasinophyceae and Dinophyceae were the dominant components of the libraries. Shi *et al.*, (2011) found that, in filtered samples, prymnesiophytes were not significant contributors to 18S rRNA gene clone libraries. However, they found that in flow cytometrically sorted samples the bias against Prymnesiophyceae disappeared and similar abundance levels were seen, compared with the use of plastid primers. In the current study prymnesiophytes were only observed in the two nuclear 18S rRNA gene clone libraries on sorted cells. This was in contrast to their abundance levels observed when using the plastid biased primers. A different set of 18S rRNA gene primers was used compared to Shi *et al.*, (2011) which could explain the lower number of nuclear prymnesiophyte sequences obtained here compared to the aforementioned study. However, the use of 18S rDNA prymnesiophyte specific primers revealed a higher diversity of prymnesiophytes in the CTD 70 sample. Using the prymnesiophyte specific primers 3 OTUs were obtained relating to *Chrysochromulina* clades B1 and B2 (order Prymnesiales) and Coccolithales. In the corresponding library created using plastid 16S rRNA primers only 2 OTUs were obtained. These were related to *Chrysochromulina* clade B2 and Coccolithales. The use of specific primers revealed a higher diversity than observed using general primers. However, the general trend that *Chrysochromulina* clade B2 was the dominant group was observed in both libraries.

Recent analysis by Shi *et al.*, (2011) has hypothesised that sequences relating to Prasinophyceae 18S clade VIIB may correspond to clade Pras 16S-VIII although, at present, no plastid sequence is available for this clade to support this hypothesis. In



this study, sequences corresponding to Pras 16S-VIII were observed in CTD 12 and 25 in the plastid biased 16S rRNA gene clone libraries. The dominant component in the corresponding CTDs in the nuclear 18S rRNA clone libraries was the 18S clade VIIB, adding further weight to the idea that these clades could correspond to each other.

Certainly, the use of different primer sets appears to give a differing view of PPE diversity in the environment. In this study, in the 18S rRNA gene libraries, Prasinophyceae was the dominant class, whilst plastid biased 16S rRNA primers found Prymnesiophyceae and Chrysophyceae to be the dominant classes. Further, Shi *et al.*, (2011) found that a different set of 16S rRNA gene primers gave a slightly different composition. This suggests that a combination of several primer sets may be required to unveil the full diversity of PPEs in the environment, or the development of novel or alternative approaches. Certainly, other approaches have already been used to assess the diversity of PPEs in the marine system. Zeidner *et al.*, (2003) undertook an analysis of photosynthetic picophytoplankton using the *psbA* gene which encodes the D1 protein of photosystem II. Their study found eukaryotic sequences related to haptophyta, stramenopiles and prasinophyta with Mamiellophyceae being especially abundant in Monterey Bay, California. A further study (Man-Aharonovich *et al.*, 2010) in the Eastern Mediterranean Sea, targeting *psbA* transcripts, found that Prymnesiophyceae sequences constituted the most numerous group, especially in surface waters, which is a similar finding to that of the plastid 16S rRNA flow cytometry results obtained in this study and in Kirkham *et al.*, (2011). Recently, 454 pyrosequencing has been used to investigate the composition and genetic diversity of picoeukaryotes in subtropical coastal waters, with primers targeting the hypervariable V4 region of the 18S rRNA gene.

Prasinophytes, especially Mamiellales, were the dominant photosynthetic component in the libraries created using 454 pyrosequencing (Cheung *et al.*, 2010). The use of 454 pyrosequencing enables high throughput analysis compared with the construction of clone libraries and could negate the use of RFLPs to detect OTUs. Gescher *et al.*, (2008) developed a DNA microarray to investigate the diversity of picoeukaryotic prasinophytes in a high throughput way. The high proportion of environmentally important prasinophytes, especially in coastal regions, which have been isolated, meant that microarray probes could be designed and tested, ensuring that the diversity in environmental samples could be assessed. This approach though may not be applicable for other classes, such as Prymnesiophyceae, where a large proportion of the sequences obtained in clone libraries bear little relation to cultured species.

#### **4.4.5 Latitudinal diversity gradient**

Since the times of early naturalists the differences in the number of coexisting organisms in environments has fascinated scientists. Among the most intriguing of phenomenon, which has interested ecologists, is the increase in diversity with a decrease in latitude. In 1807, Alexander von Humboldt wrote “The nearer we approach the tropics the greater the increase in variety of structure, grace of form and mixture of colors (sic), as also in perpetual youth and vigor of organic youth”. The latitudinal gradient in species diversity has been known since the 1800s and has been well documented in multicellular plants and animals. The evidence for an increase in species richness from the poles towards the equator is less equivocal in microbial species (Fuhrman *et al.*, 2008). A recent meta-analysis by Hillebrand (2004), which took into account 14 orders of magnitude of body mass in eukaryotes, showed that

the latitudinal gradient was observed over all size ranges. There are exceptions, such as aquatic macrophytes, which show an increase in diversity with increasing latitude (Crow, 1993). Hillebrand (2004) found that the strength of the correlation increased with increasing body mass suggesting that, at the microbial level, the latitudinal gradient may not be as obvious as is found in multicellular organisms. The lack of a conclusive pattern as seen in multicellular organisms has been attributed to high abundances, frequent and long distance dispersal and low extinction rates, although there is some evidence of latitudinal gradients in certain taxa of marine microbes, such as bacterioplankton and coccolithophorids (Barton *et al.*, 2010).

In this study, at the 55% incident light level, using the Margalef's diversity index as an estimate for species richness, a peak in species richness for the Euk-A population was observed in the equatorial region of the transect, with a decline in species richness towards the higher latitudes. A dip in species richness was observed at CTD 70 (16°S), where sequences corresponding to *Prasinoderma* (Prasinococcales) dominated the library. *Prasinoderma* is a fast growing, 'r' selected species, which may suggest that there was a perturbation in the system, such as an eddy, which led to the domination of this species at the station and thus a reduction in diversity. The data is consistent with a latitudinal diversity gradient. However, the study was restricted to 6 stations throughout the AMT and further investigation would be required to gain any statistical power to the extent of the gradient and the factors which contributed to it.

A study by Fuhrman *et al.*, (2008) investigating the possible latitudinal gradient noticed by Pommier *et al.*, (2007) in bacterioplankton, found that with increased sampling there was a negative trend in species diversity with latitude. These authors

observed that there was a significant negative correlation between diversity and latitude and a positive one with temperature which led them to suggest that the gradient may be an effect of solar radiation flux on sea surface temperature. Although the data on PPEs did not find a statistically significant trend with temperature, there was a positive trend noticed and analysis of further stations, as had been undertaken by Fuhrman *et al.*, (2008) of the Pommier *et al.*, (2007) data, may show a significant effect of temperature on PPE diversity. Fuhrman *et al.*, (2008) suggested that diversity increases with increasing environmental temperature because of the kinetics of biological processes, including rates of reproduction, dispersal, species interaction, mutation and speciation. This could be termed “the Red Queen runs faster when she is hot” hypothesis (Fuhrman *et al.*, 2008).

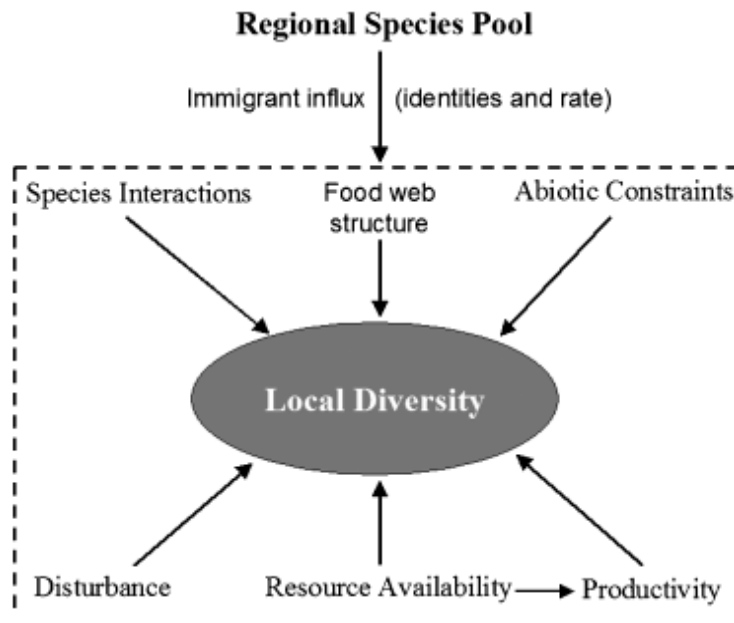
A different trend was noticed for the Euk-A PPE population at the DCM. The peak in diversity was observed at the northern gyre station (CTD 25 at 27°N). The only significant correlation was with nitrite, where a negative correlation was observed, although there were also negative correlations between diversity and phosphate and silicate. This could suggest that at the DCM maximum resource competition theory is determining the diversity of stations. Resource competition theory suggests that, in stable environments, the phytoplankton type with the lowest environmental nutrient concentration, at which death and mortality are balanced (designated as  $R^*$ ), is expected to outcompete others over time. Since  $R^*$ , for each phytoplankton type, depends on physiological characteristics and mortality, there are, in theory many different ways in which  $R^*$  can be achieved, leading to the co-existence of multiple types of phytoplankton (Barton *et al.*, 2010). Fast growing species, such as diatoms have been shown to dominate the biomass in the variable high latitude oceans. Species which are able to survive on minimal resources dominate the stable low

latitude seas (Barton *et al.*, 2010). Ptacnick *et al.*, (2008) found that resource use efficiency, in natural phytoplankton populations, was higher when the populations were more diverse. This suggests that in regions such as the oligotrophic gyres, where nutrients are limited, higher diversity may lead to higher efficiency in the use of the available nutrients. This theory could explain the high diversity present in the south Atlantic gyre at the DCM.

In a study by Cermeno *et al.*, (2008) investigating the diversity of diatoms in the Atlantic, they found no latitudinal gradient. Instead they saw a negative correlation with nutricline depth and suggested that nutrient availability in the photic zone was responsible for higher diversity in these lineages. Diatoms, by virtue of high growth rates, luxury nutrient uptake (the ability of organisms to take up more nutrients than cellular demand requires) and ability to store inorganic nutrients, can exploit unstable environments, where nutrient supply exceeds biological demand. The adaptation to unstable environments imposes severe nutritional constraints under stable conditions, seen in areas such as oligotrophic regions where their diversity is low (Cermeno *et al.*, (2008).

The data obtained here suggests that, at the 55% light level, there is a latitudinal diversity gradient whilst at the DCM, latitude seems to have less of an effect on the diversity at a site, than other factors, such as nutrient availability. The species diversity at a site is a complex set of interactions between physical, biological and chemical factors (Figure 4.22; see Smith, 2007) and further investigations, including the study of additional stations could lead to a better indication of the factors which limit the distribution of specific PPE groups. It also appears that the presence of a latitudinal gradient depends on the group studied and the position in the water

column which is investigated, with differential trends being observed for the two depths studied here.



**Figure 4.22.** The complex interactions which lead to the local diversity in a specific area (Taken from Smith, 2007).

#### **4.4 Summary**

The marine algal plastid biased 16S rRNA gene clone libraries constructed along AMT18 demonstrated a dominance of Prymnesiophyceae at the 55% light level and of Chrysophyceae at the DCM. The Prymnesiophyceae component was principally comprised of sequences related to *Chrysochromulina* clade B2 with little relatedness to species which have been isolated into culture. Prymesiohyceae were the dominant components in both flow cytometry sorted samples as well as filtered samples analysed with this approach. However, some classes were only observed in the clone libraries of one method and not in the other. Clone libraries created using

nuclear 18S rDNA primers showed a dominance of clade VII prasinophytes in the libraries. These results suggest that a combination of primers should be used to gain a complete understanding of the composition of PPEs in the marine environment. Having said this, the data suggests that that marine algal plastid biased PCR approach on sorted cells is probably the best compromise approach for assessment of PPE community structure.

CCA showed that less than 60% of the variation in PPE class and OTU distribution patterns was explained by the measured environmental variables. Further studies, investigating the impact of factors such as viral infection and grazing on the diversity of PPEs, are required in order to explain more of the variation in these AMT samples.

Diversity indexes along the transect suggest that there could be a latitudinal diversity gradient at the 55% light level although further stations would need to be analysed to increase the statistical power of the analysis. At the DCM a different pattern of diversity was observed with suggestions that the level of nutrients may play a more important role at the DCM.

## **Chapter 5**

### **Approaches to isolate novel PPEs from AMT18**



## **5.1 Introduction**

Eukaryotic phytoplankton are major contributors to primary production in marine systems. Oceanic environments cover 71% of the Earth's surface and primary production in this environment accounts for ~ 50% of global CO<sub>2</sub> fixation (Field *et al.*, 1998). Culture independent surveys of the microbial biota, that rely on amplification and sequencing of ribosomal (e.g. 16S rRNA) and conserved functional genes (e.g. *psbA*), have revealed that the microbial world is incredibly diverse. Single gene surveys are able to reveal which organisms are present in the environment but little about the metabolic abilities or functions of these organisms within the community. Studies into the physiology and morphology of microorganisms have viewed cultivation of the organism as fundamental.

The conditions for classical cultivation are far removed from the endogenous biotic and abiotic conditions found in the environment and many microbial organisms have resisted cultivation attempts (Alain and Querellou, 2009). 'Unculturability' is unlikely to be an absolute. If they are able to grow in nature then they must be capable, under a certain set of conditions, of being cultured in the laboratory. The failure is in the ability to identify the necessary conditions in which the target organisms can grow. Limited knowledge of several factors including: 1) the diversity of the organisms themselves 2) the chemistry of the natural environment in which they reside and 3) any natural biotic and abiotic interactions that exist, have led to a lack of efficient cultivation techniques (Alain and Querellou, 2009).

The ecological importance of eukaryotic algae has meant that there have been efforts to obtain into culture these organisms since the 1850's. The first published report of a cultured eukaryotic alga was in 1850 when Ferdinand Cohn managed to keep the

unicellular flagellate *Haematococcus* (Chlorophyceae) (Andersen, 2005). It was not until 1890 that the Dutch microbiologist Beijerinck obtained axenic cultures of *Chlorella* and *Scenedesmus*. At the same time Miquel, at the Montsouris Observatory (Paris, France) developed a technique using a micropipette and microscope to isolate single cells. Using this method he was able to obtain axenic cultures of several freshwater and marine diatoms. At the end of the 19<sup>th</sup> century Klebs, at the University of Basel, Switzerland, was the first person to successfully isolate algae onto agar (Andersen, 2005). In recent years new methods used in the cultivation ‘pipeline’ have been developed, including flow cytometric sorting (Vaulot *et al.*, 2008) and single cell manipulation techniques such as optical tweezers and laser microdissection (Tyson and Banfield, 2005).

Many PPEs are still known only from environmental rRNA gene sequences e.g. Prasinophyceae clade VII B (Le Gall *et al.*, 2008), whilst other groups have recently been redefined in their nomenclature e.g. the elevation of the order Mamelliales to the class Mamelliophyceae (Marin & Melkonian, 2010). Hence, knowledge of the culture conditions required for successful isolation of a novel PPE group is often limited by the lack of any cultured representative even at the class level. This is reiterated by the fact that several new algal classes with picoplanktonic representatives have been described in recent years, such as Pelagophyceae (Andersen *et al.*, 1993), Bolidophyceae (Guillou *et al.*, 1999) and Pinguiphyceae (Kawachi *et al.*, 2002). The need for cultured PPE representatives is especially acute in open ocean oligotrophic regions. Such open ocean species reside in extremely low nutrient conditions, and hence are often outgrown by fast growing “weed” species in enrichment cultures (Andersen, 2005). To date 71 species of phytoplankton have been described which have a minimum size  $\leq 3 \mu\text{m}$  (Vaulot *et al.*, 2008). Of these,

30 can be described as picoplanktonic including the “flagship” species *Ostreococcus* and *Micromonas* (both in the class Mamiellophyceae) (Vaulot *et al.*, 2008; Marin & Melkonian 2010).

The Atlantic Meridional Transect undertaken in late autumn 2008 (see chapter 4) offered a unique opportunity to attempt to obtain into culture PPEs inhabiting a range of environmental conditions, including oligotrophic gyre systems, nutrient rich equatorial regions and temperate regions. As evidenced in chapter 4 and elsewhere (e.g. see Fuller *et al.*, 2006a; McDonald *et al.*, 2007; Shi *et al.*, 2009, 2011) culture-independent studies have shown that members of the Prymnesiophyceae, Chrysophyceae and novel Prasinophyceae lineages are widespread in open ocean marine systems but with a dearth of cultured representatives. It was of particular interest then, to attempt to obtain into culture strains potentially representing members of such groups so that future physiological and biochemical studies could be undertaken.

## **5.2 Materials and methods**

### **5.2.1 Collection of PPEs for culture**

Starter cultures were collected during AMT18 (October-November 2008) using two different methods (see sections 2.1.2.2 and 2.1.2.3). At the end of the cruise, cultures were analysed by flow cytometry and cultures with potential PPEs present (as evidenced from chl *a* and side scatter flow cytometry signatures) were brought back to the UK. Tables 5.1 and 5.2 contain the cruise location details from which starter cultures were obtained using the two different methods.

**Table 5.1: Stations sampled for the collection of starter PPE cultures using direct concentration of natural seawater samples.**

Station	Depth (m)	Latitude	Longitude
45	Surface	11.8206	-32.8242
48	Surface	8.6438	-30.7085
48	66	8.6438	-30.7085
51	79	5.3398	-28.5213
56	Surface	2.791	-26.854
56	46	2.791	-26.854
56	79	2.791	-26.854
61	Surface	-6.0521	-24.978
61	87	-6.0521	-24.978
66	Surface	-12.8396	-24.9983
66	105	-12.8396	-24.9983
75	125	-19.1237	-24.9959
88	Surface	-32.1794	-29.8255
91	71	-33.7682	-32.003

### 5.2.2 Preparation of growth media used to culture PPEs

Natural seawater from the 0.1% incident light level, collected in the oligotrophic north Atlantic gyre during autumn 2009, was filtered through a 0.2 µm supor filter (Millipore) and sterilised by autoclaving at 121 °C for 15 min.

Artificial media was prepared as described in section 2.1.3 using this ‘natural’ seawater as a basis.

Where described, cultures were supplemented with a grain of rice (Uncle Ben’s) to encourage growth of heterotrophic bacteria by the slow release of organic matter. The rice was not autoclaved to prevent the release of too much organic matter (Andersen, 2005).

Where described, f/2 medium (see Table 2.2) was supplemented with silicate ( $\text{Na}_2\text{SiO}_3 \cdot 9\text{H}_2\text{O}$ ) at a concentration of 30 g L<sup>-1</sup>.

**Table 5.2: Stations sampled for the collection of starter PPE cultures using flow cytometric sorting.**

Station	Depth (m)	Latitude	Longitude
9	Surface	46.5907	-18.6966
16	129	36.0117	-27.7373
19	37	33.2979	-30.7968
19	132	33.2979	-30.7968
19	160	33.2979	-30.7968
25	Surface	27.6318	-37.0306
25	55	27.6318	-37.0306
25	129	27.6318	-37.0306
28	Surface	24.7443	-40.0885
28	173	24.7443	-40.0885
29	120	24.7443	-40.0885
32	Surface	22.5997	-40.2659
32	124	22.5997	-40.2659
38	100	16.8105	-36.2068
63	Surface	-8.8254	-24.9954
63	92	-8.8254	-24.9954
63	104	-8.8254	-24.9954
70	135	-16.6377	-24.9939
97	Surface	-37.966	-37.4111
97	50	-37.966	-37.4111
101	26	-43.2266	-44.6134

### 5.2.3 Soil extract preparation

Soil was taken from a deciduous woodland (Tocil Wood, Coventry, UK (52° 22' N, 1° 33' W) and any extraneous materials (e.g. rocks, leaves and roots) were removed. Soil was dried at room temperature and sieved to remove any larger particles present. The soil was diluted (2 parts Milli-Q water added to 1 part of soil) and autoclaved for 2 hours. Particulate matter was allowed to settle before being filtered (Whatman # 1 filter) and the extract was further autoclaved to establish a sterile solution (Andersen, 2005). The soil extract was then added to other components of this soil extract medium, as described in Table 2.2.

#### **5.2.4. Colloidal iron preparation**

Ferrihydrate was synthesized by the addition of ferric chloride ( $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$ ) to Milli-Q water to a final concentration of  $4.0 \times 10^{-4}$  M. The solution was kept at room temperature for 50 min and subsequently heated to 90 °C in a water bath for 5 min. The solution was cooled to room temperature and then added directly to growth media (see Table 2.2).

#### **5.2.5 Obtaining clonal cultures of PPEs**

Clone cultures of PPEs were obtained via agar plating. Two methods were utilised: pour plating and streaking on solid agar (see section 2.1.5).

Serial dilution was also used in an attempt to obtain clonal PPE cultures. A subsample of the starter culture (200  $\mu\text{l}$ ) was placed into a single well of a 96 well plate. Serial 1 in 10 dilutions into the appropriate fresh media were then undertaken. The plate was then maintained as described in section 2.1.3.

#### **5.2.6 Longer-term maintenance of PPE cultures**

Both starter cultures and clonal isolates were maintained over the long term as described in section 2.1.3.

#### **5.2.7 Molecular characterisation of cultures**

Nucleic acids were extracted from cultures using a chemical denaturation technique as described in section 2.2.1.1.

Characterisation of the samples was undertaken using plastid biased 16S rRNA primers (PLA491F – 1313R) using the PCR protocol detailed in section 2.2.2.1.

Restriction enzyme digestion of the DNA fragments was undertaken using the enzymes *Hae*III and *Eco*RI in a double digestion as described in section 2.2.5.

Purified PCR products (section 2.2.3) representative of unique RFLP patterns were sent to the NERC Biomolecular Analysis Facility in Edinburgh for Sanger sequencing, using an ABI 3730 instrument.

## **5.3 Results**

### **5.3.1 Isolation of PPEs enriched using soil extract medium**

Potential PPE cultures obtained following concentration of phytoplankton from natural seawater samples were sub-cultured into soil extract enriched medium (Table 2.2). The soil extract medium promoted the growth of diatom species with 7 of the 14 cultures being dominated by either *Cylindrotheca* sp. or *Diatoma tenuis* (Table 5.3 and Figure 5.2), as determined following sequencing of the plastid 16S rRNA gene. Neither of these species classify as PPEs, since cells were > 40 µm in length (as evidenced by microscopy, data not shown) but were presumably able to pass through the 5 µm pore-size filters used during concentration, since their diameter was only ~ 2 µm. The remaining strains observed in the soil extract enriched cultures were related to *Prasinoderma coloniae* (Prasinococcales) (Table 5.3 and Figure 5.1).

To remove the undesired diatom species, germanium dioxide (1 mg L<sup>-1</sup>) was added to each culture, and growth monitored over subsequent weeks. Removal of these diatom strains was successful, as evidenced from a change in RFLP pattern and subsequent 16S rRNA analysis determined that *Prasinoderma* spp. and *Chlorella* spp. strains becoming dominant in these cultures (Figure 5.1). A single culture relating to an *Ochromonas* sp. was present in sample 48\_Surface which was related

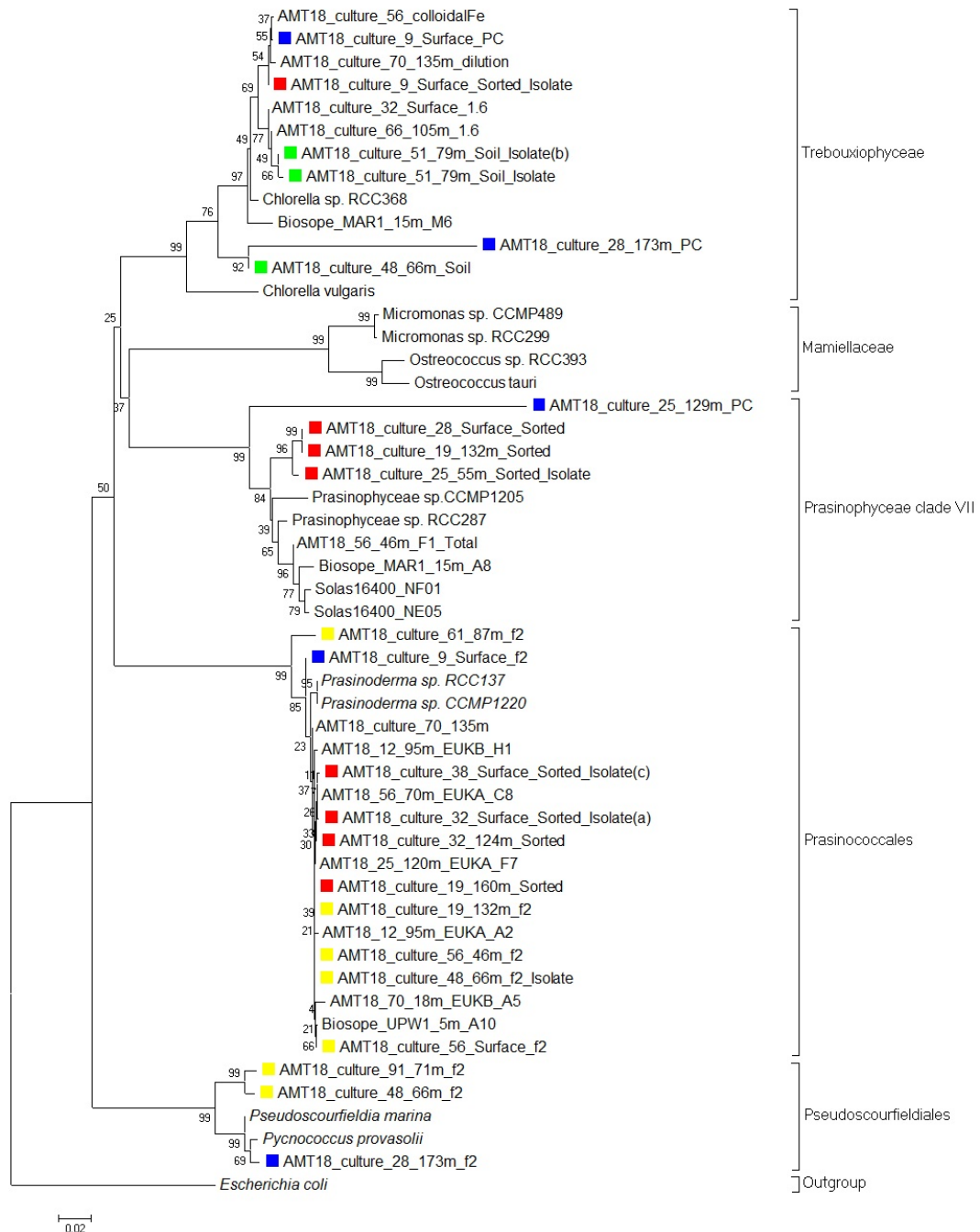
to strains observed during the AMT18 cruise (Figure 5.2 and Chapter 4). However, after treatment with germanium dioxide the *Ochromonas* sp. was no longer observed in the sample.

**Table 5.3. Cultured phytoplankton obtained from AMT 18 using soil extract medium as assessed by sequencing of the plastid 16S rRNA gene**

Station	Depth (m)	Species
45	Surface	<i>Prasinoderma</i> sp.
48	Surface	<i>Ochromonas</i> sp.
48	66	<i>Cylindrotheca</i> sp.
51	79	<i>Diatoma tenuis</i>
56	Surface	<i>Prasinoderma</i> sp.
56	46	<i>Prasinoderma</i> sp.
56	79	<i>Prasinoderma</i> sp.
61	Surface	<i>Prasinoderma</i> sp.
61	87	<i>Cylindrotheca</i> sp.
66	Surface	<i>Diatoma tenuis</i>
66	105	<i>Cylindrotheca</i> sp.
75	125	<i>Prasinoderma</i> sp.
88	Surface	<i>Prasinoderma</i> sp.
91	71	<i>Diatoma tenuis</i>

Subsequent attempts at obtaining clonal cultures using soil extract medium supplemented with agar (0.9% (w/v) final concentration), only facilitated the isolation of single *Chlorella* colonies as determined by subsequent molecular analysis (isolates from AMT18\_51\_79m\_Soil in Figure 5.1).





**Figure 5.1. Phylogenetic tree of 16S rRNA for Chlorophyta species using different culture methods. Green = Soil extract enriched cultures (Table 5.3); Yellow = f/2 enriched seawater concentrate cultures (Table 5.4); Red = Flow cytometrically sorted cultures (Table 5.5); Blue = Comparison of different culture medium (Table 5.6).**

### 5.3.2 Isolation of cultures from seawater concentrates enriched with f/2 medium

Characterisation, by the 16S rRNA sequencing of representative RFLP patterns, from seawater concentrate cultures enriched with f/2 in August 2009, showed that the majority of cultures were dominated by either of two prasinophytes. Thus, of the 15 cultures analysed, seven were dominated by strains phylogenetically related to *Pycnococcus* (Pseudoscourfieldiales), five were related to *Prasinoderma* (Prasinococcales) whilst the remaining three cultures contained a mix of species determined by RFLP analysis and were characterised by molecular techniques. Further characterisation of the f/2 enriched cultures in April 2010 showed that all cultures were dominated by the prasinophyte *Prasinoderma* sp. (Table 5.4) as evidenced by a single sequence read obtained from PCR of the plastid 16S rRNA gene.

**Table 5.4: Taxonomic identification of dominant strains in enrichment cultures from seawater concentrates supplemented with f/2 medium.**

Station	Depth	Species- August 2009	Species – April 2010
45	Surface	<i>Pycnococcus</i> sp.	<i>Prasinoderma</i> sp.
48	Surface	Mixed culture	<i>Prasinoderma</i> sp.
48	66	<i>Pycnococcus</i> sp.	<i>Prasinoderma</i> sp.
51	79	Mixed culture	<i>Prasinoderma</i> sp.
56	Surface	Mixed culture	<i>Prasinoderma</i> sp.
56	46	<i>Pycnococcus</i> sp.	<i>Prasinoderma</i> sp.
56	79	<i>Prasinoderma</i> sp.	<i>Prasinoderma</i> sp.
61	Surface	<i>Prasinoderma</i> sp.	<i>Prasinoderma</i> sp.
61	87	<i>Prasinoderma</i> sp.	<i>Prasinoderma</i> sp.
66	Surface	<i>Pycnococcus</i> sp.	<i>Prasinoderma</i> sp.
66	Surface	<i>Prasinoderma</i> sp.	<i>Prasinoderma</i> sp.
66	105	<i>Pycnococcus</i> sp.	<i>Prasinoderma</i> sp.
75	125	<i>Prasinoderma</i> sp.	<i>Prasinoderma</i> sp.
88	Surface	<i>Pycnococcus</i> sp.	<i>Prasinoderma</i> sp.
91	71	<i>Pycnococcus</i> sp.	<i>Prasinoderma</i> sp.

The f/2 enrichment cultures were separately sub cultured in f/2 medium but with dissolved iron replaced by colloidal iron in order to target mixotrophs in the cultures.

The results showed that, whereas in the f/2 enrichment the cultures were dominated by *Prasinoderma* the cultures with colloidal iron were dominated by *Chlorella*. No difference in the strains present in these subcultures was detected with *Chlorella* sp. again being the dominant organism in all cultures as represented by AMT18\_culture\_56\_colloidalFe in Figure 5.1.

### **5.3.3 Isolation of cultures enriched with artificial K medium**

Molecular characterisation of subcultures from seawater concentrates enriched with K medium was undertaken in August 2009. All the cultures were found to be dominated with strains related to *Prasinoderma*. Further molecular analysis in subsequent months using the plastid 16S rRNA gene as a means of culture assessment confirmed the dominance of *Prasinoderma* in the cultures.

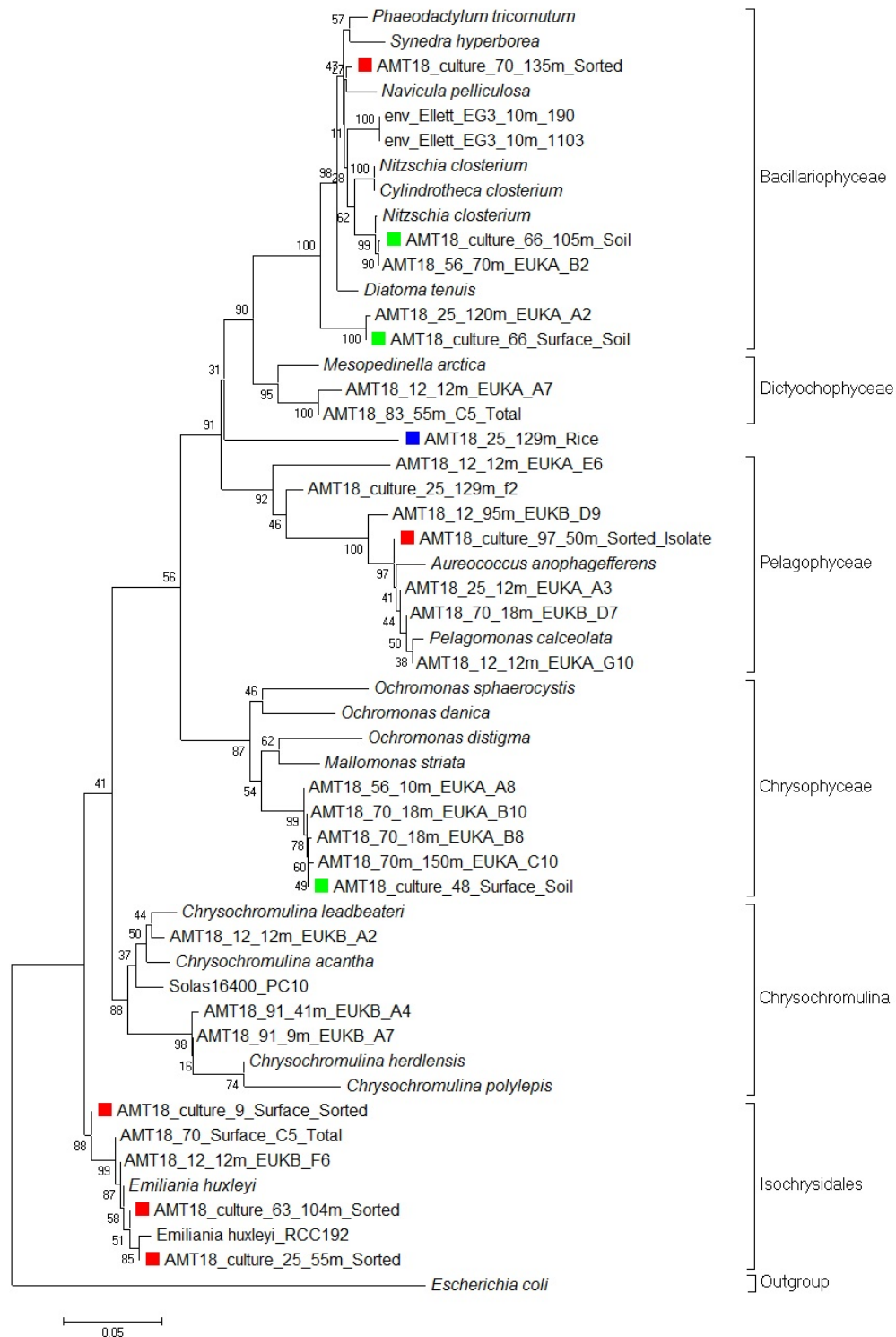
### **5.3.4 Isolation of cultures enriched with artificial f/2 medium following flow cytometric sorting of natural PPE communities obtained on AMT18**

Those enrichment cultures which resulted from flow cytometric sorting of natural PPE populations from AMT18 (see 2.1.2.2) presented a wider range of PPE diversity, again as evidenced by sequencing of plastid 16S rRNA genes. As for the seawater concentrates, the dominant class in these enrichments was Prasinophyceae, but here, whilst *Prasinoderma* was the dominant genus in many of the unsorted samples, it was only found in one enrichment (CTD 32 – Surface) in August 2009. The dominant prasinophyte (observed in six of the enrichments) during August 2009, was related to Prasinophyceae clade VIIA, whilst two further enrichments were dominated by a *Pycnococcus* sp. (Pseudoscourfieldiales) (Table 5.5 and Figure 5.1). At the time of analysis in August 2009, a further three enrichments were dominated by Isochrysiadales Prymnesiophyceae, most closely related to *Emiliania*

*huxleyi* (Isochrysiadales) as evidenced from sequence analysis (see Fig. 5.2). Subsequent molecular analysis of enrichment cultures in April 2010 showed however, that members of the Prymnesiophyceae were no longer dominant components of any of the cultures. At this time, five of the enrichments were dominated by a *Pelagomonas* sp. and two enrichments were dominated by a sequence related to *Navicula pelliculosa* as indicated by the 16S rRNA sequencing of representative RFLP patterns (Figure 5.2). Nonetheless, *Prasinoderma* sp. and Prasinophyceae clade VIIA (three cultures apiece) still dominated several of the other enrichments (see Table 5.5).

**Table 5.5. Cultured phytoplankton obtained from AMT18 following flow cytometric sorting and enrichment in f/2 medium as assessed by sequencing of the plastid 16S rRNA gene.**

Station	Depth (m)	Species - August 2009	Species - April 2010
9	Surface	<i>Emiliania huxleyi</i>	Mixed culture
16	124	<i>Pycnococcus</i> sp.	Mixed culture
19	37	Prasinophyceae clade VIIA	Mixed culture
19	132	Prasinophyceae clade VIIA	Prasinophyceae clade VIIA
19	160	Prasinophyceae clade VIIA	Mixed culture
25	Surface	Mixed culture	<i>Pelagomonas</i> sp.
25	55	<i>Emiliania huxleyi</i>	Prasinophyceae clade VIIA
25	129	Mixed culture	<i>Pelagomonas</i> sp.
28	Surface	Prasinophyceae clade VIIA	Mixed culture
28	173	-	<i>Pelagomonas</i> sp.
29	120	Prasinophyceae clade VIIA	Mixed culture
32	Surface	<i>Prasinoderma</i> sp.	<i>Prasinoderma</i> sp.
32	124	Mixed culture	<i>Prasinoderma</i> sp.
38	100	Prasinophyceae clade VIIA	Prasinophyceae clade VIIA
63	Surface	-	<i>Prasinoderma</i> sp.
63	92	-	<i>Pelagomonas</i> sp.
63	104	<i>Emiliania huxleyi</i>	Mixed culture
70	135	-	<i>Navicula pelliculosa</i>
97	Surface	-	Mixed culture
97	50	<i>Pycnococcus</i> sp.	<i>Pelagomonas</i> sp.
101	26	-	<i>Navicula pelliculosa</i>



**Figure 5.2. Phylogenetic tree of 16S rRNA for non Chlorophyta species using different culture methods. Green = Soil extract enriched cultures (Table 5.3); Yellow = f/2 enriched seawater concentrate cultures (Table 5.4); Red = Flow cytometrically sorted cultures (Table 5.5); Blue = Comparison of different culture medium (Table 5.6).**

### 5.3.5 Isolation of clonal cultures from PPE enrichments

Isolation of single colonies was undertaken using pour plates as well as plate streaking (section 2.1.5). For the enrichment cultures originating from flow cytometrically sorted cells, 88 single colonies were obtained. From CTD 25 – 55m a Prasinophyceae clade VIIA isolate was obtained four times. A single isolate phylogenetically related to *Pelagomonas* sp. was obtained from station 97–50m (AMT18\_culture\_97\_50m\_Sorted\_Isolate), whilst *Chlorella* spp. and *Prasinoderma* spp. were repeatedly isolated from several different enrichments.

Serial dilution was also undertaken as a means of obtaining clonal phytoplankton cultures, using three enrichments that originated from flow cytometric sorting (Station 9 – Surface, Station 70 – 135m and Station 101 – 26m). The surface culture from station 9 was selected in an attempt to isolate into clonal culture *E. huxleyi* from this location (Table 5.5). However, isolation attempts for this culture were unsuccessful. From station 101 – 26m, a *Prasinoderma* sp. was isolated, whilst a *Chlorella* sp. was isolated from station 70 – 135m (sequence AMT18\_culture\_70\_135m\_dilution in Figure 5.1).

### 5.3.6 Isolation of PPEs into culture: a comparison of different media

A subset of the flow cytometry sorted enrichment cultures were used to undertake a comparison of four different medium types (f/2 medium, f/2 medium + silicate, PC medium and 0.2 µm filtered deep Atlantic water supplemented with rice – see Table 2.2) for potential growth of PPEs.

An f/2 medium enrichment from station 25-129m produced a strain related to *Aureococcus* sp. (Pelagophyceae) as the dominant component in the culture.

However, in f/2 + silicate and PC medium a Prasinophyceae clade VIIA strain dominated the cultures, whilst in the rice supplemented natural seawater medium an uncultured dictyochophyte species was present (Table 5.6).

Differences in composition were also observed in station 28 – 173m, where a strain related to *Pycnococcus* was observed in both the f/2 and f/2 + silicate media, while *Chlorella* was present in the PC media. No PPEs were observed in the natural seawater culture (Table 5.6).

*Chlorella* was also the dominant strain in the PC media at station 9 – Surface, while strains relating to *Prasinoderma coloniae* were observed in the f/2 and f/2 + silicate cultures (Table 5.6).

**Table 5.6: Effect of medium type on the composition of PPE enrichment cultures from AMT18**

Sample location	Medium type			
	f/2	f/2 + silicate	PC	SW + Rice
9 – surface	<i>Prasinoderma</i> sp.	<i>Prasinoderma</i> sp.	<i>Chlorella</i> sp.	-
25 - 129m	<i>Aureococcus</i> sp.	Prasinophyceae clade VII	Prasinophyceae clade VII	Uncultured Dictyochophyceae
28 - 173m	<i>Pycnococcus</i> sp.	<i>Pycnococcus</i> sp.	<i>Chlorella</i> sp.	-

## **5.4 Discussion**

In the current study enrichment cultures using soil extract medium were dominated by diatom species, which often hamper attempts to culture other algae (Andersen, 2005). The presence of the two diatom species (*Diatoma tenuis* and *Cylindrotheca* sp.) was undesirable as neither could be classified as PPEs with a length of >40 µm. To remove diatoms from the enrichment cultures, germanium dioxide (1 mg l<sup>-1</sup>) was added, since this chemical can replace silica in biochemical reactions with lethal

effect (Andersen, 2005). However, even at high concentrations, germanium dioxide appears not to affect most other algal classes e.g. Chlorophyceae, Rhodophyceae and Phaeophyceae (Lewin and Reimann, 1969). An exception is the Chrysophyceae, whose inhibition by germanium dioxide may explain the loss of the putative *Ochromonas* sp. observed in the surface culture from station 48 (Table 5.3).

Enrichment cultures have long been used as a preliminary step towards single cell isolation. The addition of soil extract is useful when the nutrient requirements of the desired alga are unknown (Andersen, 2005). The success of the technique is dependent on the soil chosen. Undisturbed deciduous woodland soils, such as those from Tocil Wood used here, appear to be preferred, especially if they have little exposure to chemicals (Tompkins *et al.*, 1995). However, a study of freshwater chrysophytes demonstrated that only two of 40 soil types were reliable for the maintenance of these algae (Tompkins *et al.*, 1995). Hence, future PPE isolations using soil extract medium should attempt to obtain soil extract from an array of different soil types.

Flow cytometric sorting of PPEs prior to enrichment in various culture media (section 5.3.4) produced a larger taxonomic diversity of phytoplankton cultures, than those obtained following filtration. This is in accordance with the data obtained by Le Gall *et al.*, (2008) in attempts to isolate PPEs from the BIOSOPE cruise in the south east Pacific Ocean. Indeed, these authors also obtained very similar taxa into culture as those isolated here. The flow sorting procedure used here facilitated between 5000 and 50,000 PPE cells being sorted per starter culture. However, robotic automation of flow cytometry does permit single cell sorting to be undertaken. High purity cultures can be obtained using single cell sorting but there is a lower level of growth success compared with larger numbers of cells in the starter



culture. Thus, Crosbie *et al.*, (2003) targeted picocyanobacterial strains from oligotrophic and oligo-mesotrophic subalpine lakes, using single cell sorting. The technique allowed the rapid establishment of a diverse collection of picocyanobacteria with a high degree of genetic purity in the cultures. Despite a low level of sort recovery and growth success, the use of single cell sorting and robotic automation means that the labour intensive isolation step of classical culturing can be bypassed, meaning a higher throughput of cultures can be subsequently analysed (Crosbie *et al.*, 2003). A combination of single cell sorting, with the use of different culture media, especially low nutrient media, could enhance the ability to isolate novel PPE strains, especially from the low nutrient oligotrophic regions which have so far been more difficult to culture, and should be strongly considered in future isolation attempts.

Among the Chlorophyta strains isolated on AMT18, several were related to *Prasinoderma* sp., which has been isolated from a range of environments (Fuller *et al.*, 2006a; Le Gall *et al.*, 2008). Moreover, four strains belonging to Prasinophyceae clade VIIA were isolated from station 25 -55m located in the North Atlantic gyre (Table 5.5). Members of this group await formal classification, and given their presence in nuclear clone libraries from the northern gyre, northern temperate and upwelling regions along AMT18 (section 4.3.6) further characterisation of these isolates is desirable. Unfortunately, whilst Prasinophyceae clade VIIA sequences were observed in flow cytometry sorted starter cultures from other stations, attempts to isolate them into culture were unsuccessful. The four cultures obtained were derived from a variety of depths, from the surface down to 160m in the northern gyre (station 19), which is a similar distribution to that found by Le Gall *et al.*, (2008) who isolated strains from both surface waters and at 100m.

Only four different isolates were obtained using agar plating. Both pour plating and streaking were used to isolate PPEs, with no differences being observed in the types of strains isolated (section 5.3.5). The majority of the strains isolated were *Prasinoderma* sp. or *Chlorella* sp. Both these species are fast growing and cosmopolitan, which is consistent with their exploitation of the high nutrient levels in the artificial media used here. In nutrient rich artificial media, community members with ‘r’ strategies often over grow and outcompete the naturally abundant ‘K’ strategists (Alain and Querellou, 2009). Watve *et al.*, (2000) found that colony counts for bacteria, in a number of environments, were higher when using dilute culture media compared to the rich nutrient media normally used. The use of dilute culture media meant that these authors were able to isolate novel ‘k’ strategists which had been outcompeted in the higher nutrient conditions. The future use of much lower nutrient media than used here, would potentially enable PPEs naturally occurring in the oligotrophic gyres to be selected for (Watve *et al.*, 2000), though this would assume that there are no ‘unusual’ nutrient requirements of such organisms, whilst also producing only low-biomass enrichments that ultimately may limit subsequent physiological and biochemical analysis.

Photosynthetic organisms have a high requirement for cellular iron and its acquisition involves a complex set of extracellular reactions (Nodwell and Price, 2001). The replacement of soluble iron chloride with colloidal iron (section 5.3.2), was intended to select for mixotrophic organisms which could acquire particulate forms of iron by ingestion. Colloidal forms of iron possess a low solubility and have, until recently, been thought to be unavailable to phytoplankton (Nodwell and Price, 2001). The mixotrophic chrysophyte *Ochromonas* could ingest particulate iron in the form of ferrihydrate with the suggestion that this uptake of particulate iron could

give mixotrophs a selective advantage in low nutrient areas (Nodwell and Price, 2001). Given that some species of *Chlorella* have a mixotrophic lifestyle (Scarsella *et al.*, 2009), this could explain the dominance of this genus in cultures grown with colloidal iron (section 5.3.2).

## **5.5 Conclusions and future directions**

Whilst most of the phytoplankton strains isolated during this study were represented in the clone libraries constructed along AMT18 (chapter 4), most are likely fast growing ‘r’ selected species. Thus, no isolates representative of the plethora of *Chrysochromulina* or environmental chrysophyte sequences found in the oligotrophic gyres along AMT18 (chapter 4) were obtained. Clearly, then there is still a significant gap between the extensive diversity of PPE sequences found in environmental surveys compared to the number of PPEs isolated into culture. The use of classical cultivation methods have led to a bias towards the fast growing ‘r’ selected species, whereas the ‘k’ selected species, which are environmentally important in areas such as the oligotrophic gyres, have so far been resistant to cultivation attempts (Andersen, 2005).

It is possible that an integration of knowledge, from the cell to the ecosystem, will be required, in order to further efforts into the cultivation of PPEs. An increase in the knowledge of the desired species through both metagenomic techniques and culture studies, will lead to a better understanding of how these species interact with their environment. With greater knowledge of these interactions, the cultivation techniques can be optimised towards the cultivation of novel organisms.

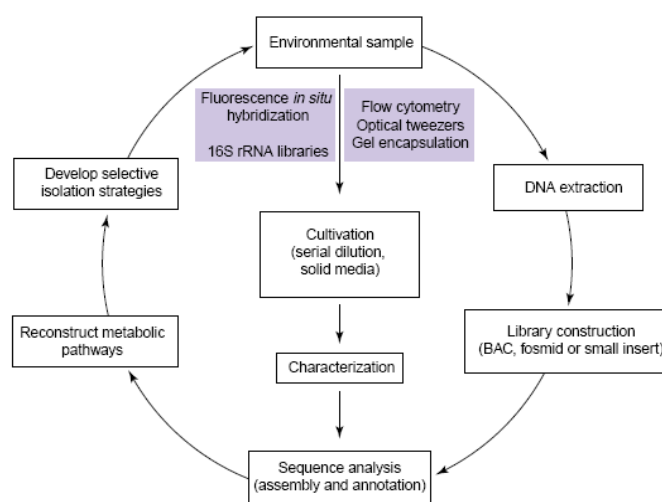
The ability to isolate novel organisms in the future will certainly require a greater knowledge of the biotic and abiotic factors which control the distribution of PPEs *in*

*situ*, as well as the use of new culturing techniques such as those currently being applied to the cultivation of bacterioplankton (see Alain and Querellou, 2009 for a review).

One technique successfully used to isolate marine bacterioplankton into culture utilises a combination of flow cytometry and single cell encapsulation (Zengler *et al.*, 2002). Cell agarose mixtures are emulsified producing gel microdroplets (GMD), and then grown in culture medium for at least five weeks for the formation of colonies to occur in the GMDs. Since all the encapsulated cells are cultivated together and the gel matrix is large, metabolites and other molecules, such as signalling molecules, are allowed to freely pass in the media, more accurately simulating environmental conditions (Zengler *et al.*, 2002). GMD culturing is performed in an open continuously fed system which does not allow the build up of toxic end products, thus more closely replicating the open ocean conditions found in natural environments. GMD-containing colonies could then be distinguished from empty GMDs, or free living cells and could be sorted into 96 well microtiter plates for further characterisation. However, the isolation of GMDs into microtiter plates disrupts the cell to cell interactions which were present before isolation. Alain and Querellou (2009) proposed a modification of this method, where GMDs are sorted into microbioreactors, directly linking them to the GMD culture community, and keeping cell-cell interactions intact. This production of mixed microbial communities offers further potential to ‘cultivate the uncultivated’. Certainly, it can be expected that in the natural environment many organisms will live as part of a inter-linked ‘community’, with cells exchanging dedicated signalling molecules, trading metabolites and competing for limited resources (West *et al.*, 2007). Mixed microbial assemblages also have the capacity to perform multi-step functions that are

not possible for individual species. In some cases cooperation within the microbial community involves the sharing of metabolic intermediates or micronutrients. When these dependencies are identified they can be reproduced experimentally allowing for the growth of the desired organism in monoculture (Alain and Querellou, 2009). In the current study a higher diversity of strains was identified in the mixed start cultures than could be subsequently isolated (e.g. *E. huxleyi*). This could be due simply to the limitations of growing PPEs on agar plates, as not all algae are able to grow on plates (Andersen, 2005), but could clearly also be due to their dependencies on shared products which would not exist in isolation.

Further evidence of the importance of signalling molecules to the cultivation of marine bacterioplankton was obtained by Bruns *et al.*, (2002). Although the addition of signalling molecules, to improve cultivation, has not been explored systematically it was found that marine bacterioplankton have high affinity uptake systems for cyclic AMP (cAMP). In laboratory cultures, which have been starved of nutrients, the addition of cAMP has been shown to prevent substrate accelerated death (Bruns *et al.*, 2002). The addition of cAMP, which is part of the *crp* activation system in enteric bacteria, enables the cultivation of novel strains of bacterioplankton which were at a low concentration in the community (Bruns *et al.*, 2002). The use of signalling molecules and shared metabolic intermediaries has not been undertaken with PPEs but the approach is a possible way in which novel algae can be isolated into culture.



**Figure 5.3. Flow diagram showing possible integrations of techniques to further understand environmentally important species (Taken from Tyson and Banfield, 2005)**

As community genomic databases increase, the possibility of bypassing bottlenecks in cultivation may also be possible. With the characterisation of cultures which are presently available, combined with analysis of community genomics, reconstruction of metabolic pathways, important in the environment can be achieved. From knowledge of the specific metabolic pathways active in the environment, selective isolation strategies can be identified, to enable isolation of previously uncultured organisms (Figure 5.3). Tyson *et al.*, (2004), undertook a random shotgun sequence analysis of a low diversity acid mine drainage system, to reconstruct genomes of the bacteria and Archaea present in the system. In acid mine drainage nitrogen fixation is an essential process and genome analysis found that the only nitrogen fixation operon present in the environment was assigned to *Leptospirillum* group III. Subsequently, a diluted environmental sample was inoculated into nitrogen-free media, resulting in the growth of only the target organisms (Tyson and Banfield, 2005). This study highlights the possibility of how deductions in functional partitioning, between community members, can aid in the ability to culture novel

organisms. As the number of PPE genomic studies increases, this approach could be utilised in the future to obtain novel cultures.

Further attempts at culturing novel PPEs will certainly need to be undertaken, in order that physiological studies can be undertaken to reveal the metabolic capabilities of PPEs, and thus a greater understanding of biogeochemical cycles in the ocean. At present, failures in our ability to do so can only potentially be met using metatranscriptomic approaches (see chapters 6 and 7).

## **Chapter 6**

### **Transcriptomics of cultured marine PPEs focusing on a clade VII prasinophyte and a marine *Ochromonas* sp. (Chrysophyceae)**



## **6.1 Introduction**

Marine photosynthesis is dominated by planktonic organisms, comprised of prokaryotes and unicellular algae. The genomic analysis of the prokaryotic component of the phytoplankton is further advanced than that of the eukaryotic. In the prokaryotic component at least 23 genomes have been sequenced belonging to the dominant cyanobacterial genera *Prochlorococcus* and *Synechococcus* (see e.g. Scanlan *et al.*, 2009). The documentation of eukaryotic algal genomics first began with the publication of the nucleomorph genome of the cryptomonad *Guillardia theta* (Douglas *et al.*, 2001). Since then, genomes have been completed for two diatoms: *Thalassiosira pseudonana* (Armbrust *et al.*, 2004) and *Phaeodactylum tricornutum* (Bowler *et al.*, 2008), and the haptophyte *Emiliania huxleyi* (see Worden and Allen 2010). Alongside these genomes, 5 genomes have been published for organisms which classify as PPEs. Amongst these, four of these sequenced genomes belong to two genera in the Mamiellophyceae, *Ostreococcus* and *Micromonas* (Derelle *et al.*, 2006; Palenik *et al.*, 2007; Worden *et al.*, 2009) whilst the other is from the pelagophyte *Aureococcus anophagefferens* (Gobler *et al.*, 2011).

Genome analysis of marine phytoplankton has revealed different strategies to overcome carbon limitation in the aquatic system (Reinfelder, 2011). Photorespiration (C2 cycle) has been shown to occur in all known plants and algae highlighting its importance in the recycling of carbon. Analysis of the diatom genomes has also identified a complete urea cycle (Armbrust *et al.*, 2004). Whilst the function of the urea cycle in diatoms is far from being

completely understood a possible role in photorespiration has been hypothesised (Allen *et al.*, 2006). The hypothesis suggests the presence of a urea cycle and its connection to the recycling of carbon and nitrogen may reflect an adaptation to enhance the efficiency of carbon fixation. Most marine algae employ carbon concentrating mechanisms (Raven *et al.*, 2008). These mechanisms concentrate CO<sub>2</sub> in the vicinity of the RuBisCO active site. Carbonic anhydrases (CAs) allow for the interconversion of HCO<sub>3</sub> and CO<sub>2</sub>, and genome analysis has shown a different composition of carbonic anhydrases (CA) between species (Parker *et al.*, 2008). In the freshwater alga *Chlamydomonas*, 10 putative CAs have been identified representing  $\alpha$ ,  $\beta$  and  $\gamma$  types (Spalding, 2008) whilst the marine *Ostreococcus* only contains a  $\beta$  CA (Derelle *et al.*, 2006). The genomes of the two showed that there was surprisingly little similarity in CA which has been suggested to be an adaptation to different ecological niches (Kroth *et al.*, 2008). In the prymnesiophyte *E. huxleyi*, a  $\gamma$  CA was hypothesised to be involved in the process of calcification whilst a  $\delta$  CA was believed to be involved in carbon acquisition (Soto *et al.*, 2006). One of the major revelations revealed by genome sequencing was the identification of a putative C<sub>4</sub> photosynthetic system in unicellular organisms (Armbrust *et al.*, 2004). Different mechanisms have been proposed for C<sub>4</sub> photosynthesis in the diatom *T. pseudonana* and prasinophytes with further investigations needed to confirm these genomic observations. Diatoms are proposed to concentrate CO<sub>2</sub> in the mitochondrion before transport to the chloroplast whilst the proposed pathway for *Ostreococcus* is based on the

decarboxylation of malate in the chloroplast (Kroth *et al.* 2008; Worden *et al.*, 2009).

In the marine environment, where nutrients are frequently limiting to growth rate and/or yield, resource acquisition is critical for survival. Nitrogen is essential for life and marine algae can utilise inorganic ( $\text{NO}_3$ ,  $\text{NO}_2$  and  $\text{NH}_4$ ) and organic (urea and amino acids) forms of nitrogen (e.g. Armbrust *et al.*, 2004; Derelle *et al.*, 2006). Symbiotic relationships with diazotrophs can also be a source of nitrogen especially in diatoms (Foster *et al.*, 2011). Diatoms are found in nutrient rich waters where they can effectively out-compete other phytoplankton (Parker *et al.*, 2008). The sequencing of two diatom genomes has shown that multiple transporters for nitrate and ammonium are encoded in the genome in addition to transporters for urea and other forms of organic nitrogen (Armbrust *et al.*, 2004). This wide range of transporters likely reflects the variety of nitrogen sources that are available in the nutrient ‘rich’ habitats where diatoms dominate. In contrast, the *Ostreococcus* genome shows a different range of nitrogen acquisition genes (Derelle *et al.*, 2006). Whilst there are only single copies of several genes involved in nitrogen assimilation, there are 4 putative ammonium transporters suggesting that these prasinophytes may be strong competitors in more open ocean regions where ammonium drives production.

Genomic analysis so far undertaken on marine algae has thus revealed metabolic pathways which had previously not been described in the organisms studied. To gain further understanding of the metabolic functioning of specific PPE groups that have thus far been poorly

characterised by molecular approaches, as well as to provide a database of expressed sequence tags that could be used to aid assignment of environmental transcripts (see chapter 7), a transcriptomic analysis was undertaken on two environmentally relevant PPEs. The environmental relevance of the chosen organisms was as particularly evidenced by their prevalence in 18S rRNA or 16S rRNA clone libraries on AMT18. The targeted PPEs included a clade VIIA prasinophyte (based on 18S rRNA taxonomy), strain RCC1124, a group that appears widespread in both the Atlantic (see chapter 4) and Pacific Oceans (Shi *et al.*, 2009), and a marine *Ochromonas* sp. (strain CCMP584) a member of the Chrysophyceae, a class that is particularly abundant in plastid 16S rRNA based libraries (Fuller *et al.*, 2006a; McDonald *et al.*, 2007; Lepère *et al.*, 2009; Shi *et al.*, 2011; see chapter 4). Chrysophytes were also a significant component of the AMT18 clone libraries. A transcriptomic approach to analysing these cultures would further our knowledge of the functional abilities of these organisms.

## **6.2 Materials and Methods**

### **6.2.1 Cultures used for transcriptomic analysis**

The *Ochromonas* sp. (CCMP584) culture was isolated from the Sargasso Sea (35°N, 65° W) in 1980. The culture was maintained as described in section 2.3.1.1.

The clade VII prasinophyte (RCC1124) was isolated in the north Atlantic (48°50' N, 16°29'W) by L. Jardillier in 2006. The culture was maintained as described in section 2.3.1.2.

### **6.2.2 Methods for RNA extraction**

Two methods of RNA extraction were compared during the study with the preferred method being used in subsequent analysis.

Method 1 utilised Trizol™ (Invitrogen) for the extraction. Cells were centrifuged and resuspended in 1.5 ml of Trizol™. The suspension was left at room temperature for 10 min before 0.3 ml of CHCl<sub>3</sub> was added and shaken for 30 sec. A further 10 min incubation at room temperature was undertaken followed by centrifugation at 16,000 x g for 10 min. The aqueous layer was removed and 0.5 vol isopropanol was added and the solution was incubated on ice for 12 min. The solution was centrifuged at 16,000 x g for 10 min and the RNA pellet was subsequently washed in 75% (v/v) ethanol and air dried.

The second method utilised a Ribolysing method adapted from Gilbert *et al.*, (2000). This method of RNA extraction is described in section 2.3.2.

### **6.2.3 Enrichment of mRNA from the total RNA**

Eukaryotic mRNA was enriched in the samples using poly (dT) magnetic beads (NEB). The full method is described in section 2.3.3. Two rounds of enrichment were undertaken to minimise the amount of contamination from bacterial RNA and eukaryotic rRNA.

### **6.2.4 Construction of cDNA libraries from enriched eukaryotic mRNA**

The enriched mRNA was used as a template for the construction of cDNA libraries using Invitrogen's CloneMiner II cDNA library construction kit.

The library was constructed as described in section 2.3.5. Plasmids were extracted from the clones using Qiagen's Miniprep kit. Plasmids were digested using the restriction enzyme *Bsr*GI (NEB) and inserts were size fractionated between 250bp – 2kb (section 2.3.5).

#### **6.2.5 Sequencing of cDNA**

Size fractionated inserts were pooled together and 5 µg of cDNA was sent to the NERC Biomolecular Analysis Facility in Liverpool. The inserts were sequenced on a Next Generation Roche 454 FLX sequencer. Data was outputted as raw reads with a corresponding quality file for the sequencing. Reads were also assembled into contigs using the Newbler assembler software (Roche).

#### **6.2.6 Phylogenetic analysis of cDNA contigs**

The contigs were analysed using BLASTx against the NCBI nr database. Phylogenetic analysis of the top hits was undertaken using MEGAN as detailed in section 2.4.3.6.

Contigs relating to rRNA were analysed using BLASTn against the NCBI nr database (section 2.4.3.2)

#### **6.2.7 Functional prediction of cDNA reads**

Analysis of the contigs was undertaken using workflows from the CAMERA website.

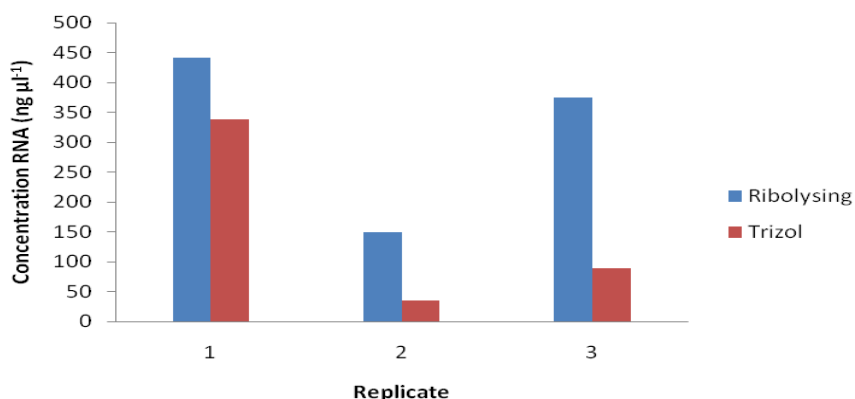
The Rapid Analysis of Multiple Metagenomes with a Clustering and Annotation Pipeline (RAMMCAP) was used to predict rRNA, identify

ORFs and analyse the predicted ORFs against protein databases such as PFAM and TIGRFAM as detailed in section 2.4.3.3. Further analysis was undertaken using workflows from CAMERA with functional prediction against the databases Cluster of orthologous groups for eukaryotic genomes (KOG) (section 2.4.3.4) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (section 2.4.3.5).

## **6.3 Results**

### **6.3.1 Comparison between RNA extraction methods**

Replicates of both methods were undertaken on a prasinophyte culture to analyse the reproducibility of the two methods. In general the Ribolysing technique was found to produce a greater quantity of RNA (Figure 6.1) and of a higher quality than the Trizol method (data not shown). Variation in RNA concentration was due to different concentrations of cells being used for the extraction. This led to the Ribolysing method to be used for the extraction of total RNA for the two transcriptomes described here.

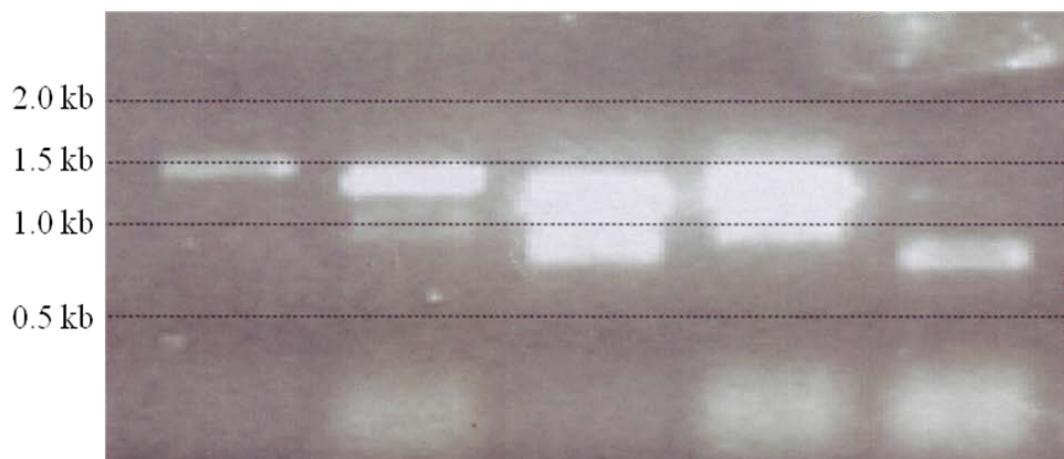


**Figure 6.1. Comparison of the concentration of RNA using two methods of RNA extraction**

### 6.3.2 Construction of cDNA libraries

Prior to library construction 18S rRNA analysis of the cultures was undertaken to determine the phylogeny of the cultures. 18S rRNA analysis of the clade VII prasinophyte was closely related to the clade VIIA prasinophyte RCC287 (Figure 6.5). 18S rRNA analysis of the *Ochromonas* sample had a top hit against *Ochromonas* CCMP584.

On average ~ 3.5 µg of enriched mRNA was obtained from 100 ml of the *Ochromonas* sample after total RNA extraction and mRNA enrichment. This was then used for library construction. After library construction, ~ 10,000 clones were picked ( in batches of ~200) and inserts extracted and size fractionated between 250 bp – 2 kb (Figure 6.2).



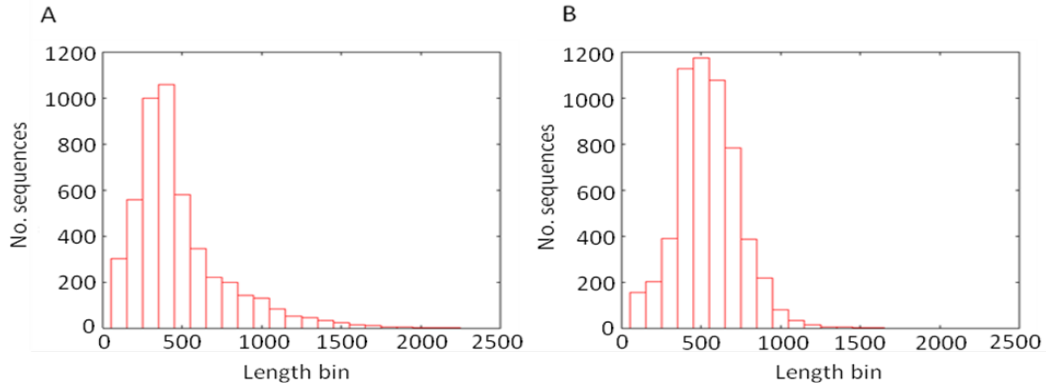
**Figure 6.2. cDNA insert size from the prasinophyte clade VIIA transcriptome sample. Each lane represents an insert from a unique clone.**

### 6.3.3 Sequencing results

High throughput sequencing of cDNAs obtained from the clade VIIA prasinophyte, strain RCC1124, produced 172,518 reads. The reads ranged in length between 30-784 bp with an average read length of 302 bp. Using a meta assembly program in CAMERA (see section 2.4.3.1), 4840 contigs

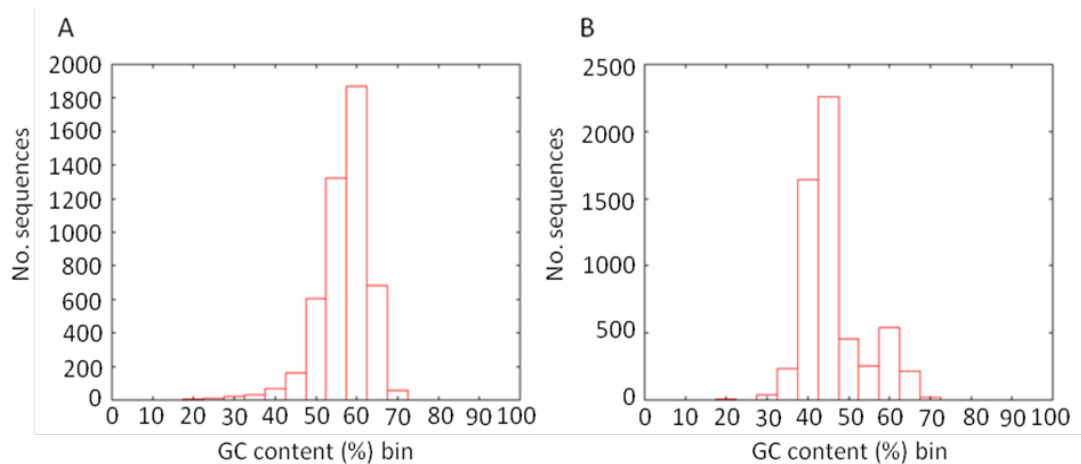


were produced with an average length of 531bp (Figure 6.3A). The modal binned GC content was 60% (Figure 6.4A).



**Figure 6.3. Distribution of contig lengths in A) Prasinophyte clade VIIA strain RCC1124 and B) *Ochromonas* sp. strain CCMP584.**

Similar high-throughput sequencing of cDNAs obtained from the *Ochromonas* sp. strain CCMP584 resulted in 206,760 reads being produced. The average length of the reads was 345bp with a range between 24-686 bp. The Newbler assembly program was used to assemble the reads into contigs. A total of 5667 contigs were produced. The average length of the *Ochromonas* contigs was 593 bp (Figure 6.3B). The GC content of the *Ochromonas* reads showed a bimodal distribution with one peak being at 45% with the second smaller peak at 60% (Figure 6.4B).



**Figure 6.4: The binned GC content (%) of A) Prasinophyte clade VIIA strain RCC1124 and B) *Ochromonas* sp. strain CCMP584.**

#### 6.3.4 Phylogenetic analysis of transcriptomes

Poly (dT) magnetic beads were used to selectively target eukaryotic mRNA and to reduce the contribution in each sample from rRNAs (see section 2.3.3). Thus, of the 4840 contigs obtained for the clade VIIA prasinophyte only 16 (0.33 %) contained rRNAs. Phylogenetic analysis (of those contigs containing reads associated with the 18S rRNA gene) against the NCBI nr database showed that the contigs were 98% identical to another clade VIIA prasinophyte isolate, strain RCC287, as well as environmental sequences obtained from the Pacific Ocean (see Fig. 6.5). BLASTx analysis was undertaken on the contigs. Phylogenetic analysis of the top hits showed that 2163 (44.7%) contigs were assigned to Eukaryota. Only 228 were assigned to Bacteria (4.7%) with the remainder unable to be assigned. Of the sequences assigned to Eukaryota the majority (70.6% (31.6% of all contigs)) were assigned to Viridiplantae with a small percentage related to Fungi/Metazoa (2.9% (1.3% of all contigs)) and stramenopiles (1.6% (0.7%

of all contigs)). Within the Viridiplantae, Chlorophyta were the dominant group, especially the prasinophytes *Micromonas* and *Ostreococcus* (see Figure 6.6).

In the *Ochromonas* sp. CCMP584 sample, of the 5667 contigs, 70 contained rRNAs (1.2%). 18S rRNA reads phylogenetically related to *Ochromonas* sp. CCMP584 (99%) were observed in the sample (see Table 6.1). However, there were also rRNA reads phylogenetically related to Trebouxiophyceae (100% identity) and a Cercozoan heterotroph (99% identity). Analysis of large subunit (28S) rRNA transcripts showed them being related to Chrysophyceae, Trebouxiophyceae and a Cercozoan. Phylogenetic relationships were lower than those observed for the 18S rRNA (Table 6.2). BLASTx analysis was again undertaken on the contigs and subsequent phylogenetic analysis showed that the majority of the top hits were related to eukaryotes (62.5%). Only 3.1% of the top hits were related to bacteria with the remainder of the contigs unassigned. Within the Eukaryotes the majority of sequences were related to Chlorophyta, especially the trebouxiophyte, *Chlorella* (29.0% of eukaryotic contigs (18.1% of all contigs)). A smaller percentage of the Eukaryotic contigs were related to stramenopiles with only 0.2% related to Chrysophyceae (0.12% of all contigs) (Figure 6.7).

**Table 6.1. Phylogenetic analysis of 18S rRNA reads from the *Ochromonas* sp. CCMP584 transcriptome**

Contig	Phylogeny	% identity
619	<i>Ochromonas</i> CCMP584	99
808	<i>Massisteria</i> sp.	99
1167	<i>Nannochloris</i> sp.	100
1658	<i>Micractinium</i> sp.	99
3889	<i>Massisteria</i> sp.	99
5011	<i>Nannochloris</i> sp.	100
5224	<i>Nannochloris atomus</i>	100
5242	<i>Nannochloris</i> sp.	99
5509	Uncultured eukaryote	99
5531	<i>Ochromonas</i> sp. CCMP584	100
5570	<i>Nannochloris</i> sp.	100
5587	<i>Ochromonas</i> sp. CCMP584	98

### 6.3.5 Classification of ORFs against the KOG database

ORF prediction using the Metagene prediction method in CAMERA was undertaken. In the clade VIIA prasinophyte cDNAs 5519 ORFs were predicted. The average length of each ORF in the sample was 112 aa. For the *Ochromonas* sp. cDNAs a similar number of ORFs (5398) were predicted with an average length of 119aa. ORFs were compared against the KOG (EuKaryotic Orthologous Groups (KOG) is a eukaryote-specific version of the Clusters of Orthologous Groups (COG) tool for identifying ortholog and paralog proteins), PFAM (protein families) and TIGRfam (a resource consisting of curated multiple sequence alignments, Hidden Markov Models (HMMs) for protein sequence classification) databases (see section 2.4.3.3 and 2.4.3.4). In the prasinophyte sample there were 2290 hits (41.5%) against the KOG database. In the *Ochromonas* sample there were 2343 hits against the KOG database accounting for 43.4% of the predicted

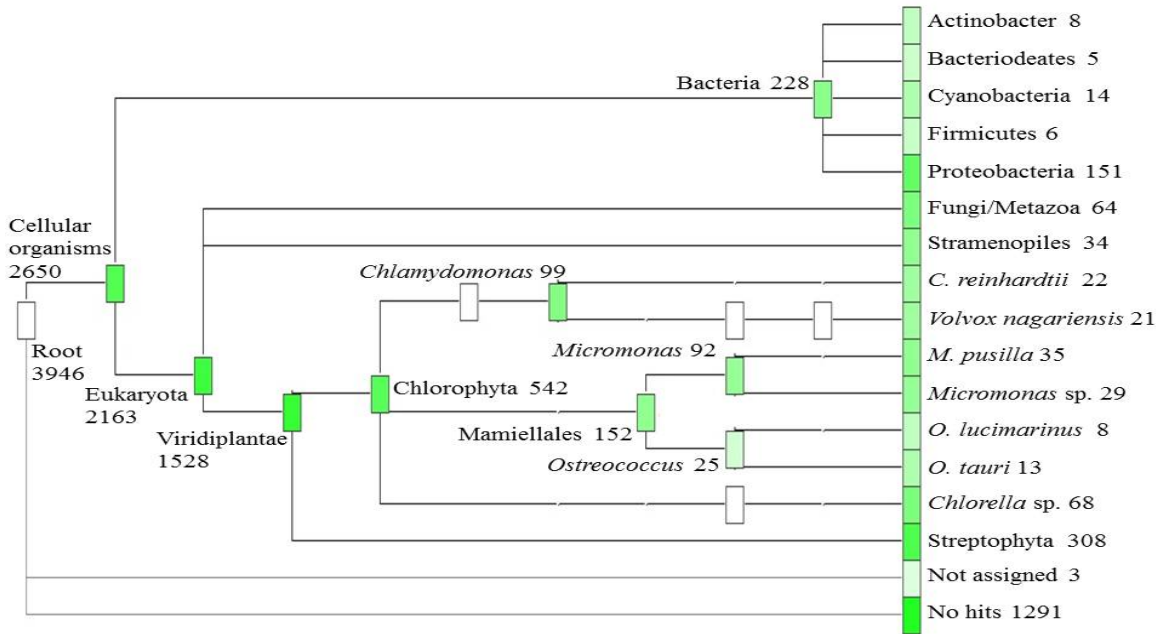
ORFs. The KOG classes with the highest percentage of hits in both samples were class O (post-translational modification, protein turnover, chaperones) and class J (translation, ribosomal structure and biogenesis). A significant fraction of the hits (9.3% (3.8% against total ORFs) and 8.7% (3.8% against total ORFs) for the prasinophyte and *Ochromonas* samples, respectively) could only be annotated as general function (class R). Classes C (Energy production and conversion), E (Amino Acid transport and metabolism) and were also significant contributors in the KOG classification in both samples (Figure 6.8).

**Table 6.2. Phylogenetic relationship of 28S rRNA transcripts in the *Ochromonas* sp. CCMP584 transcriptome.**

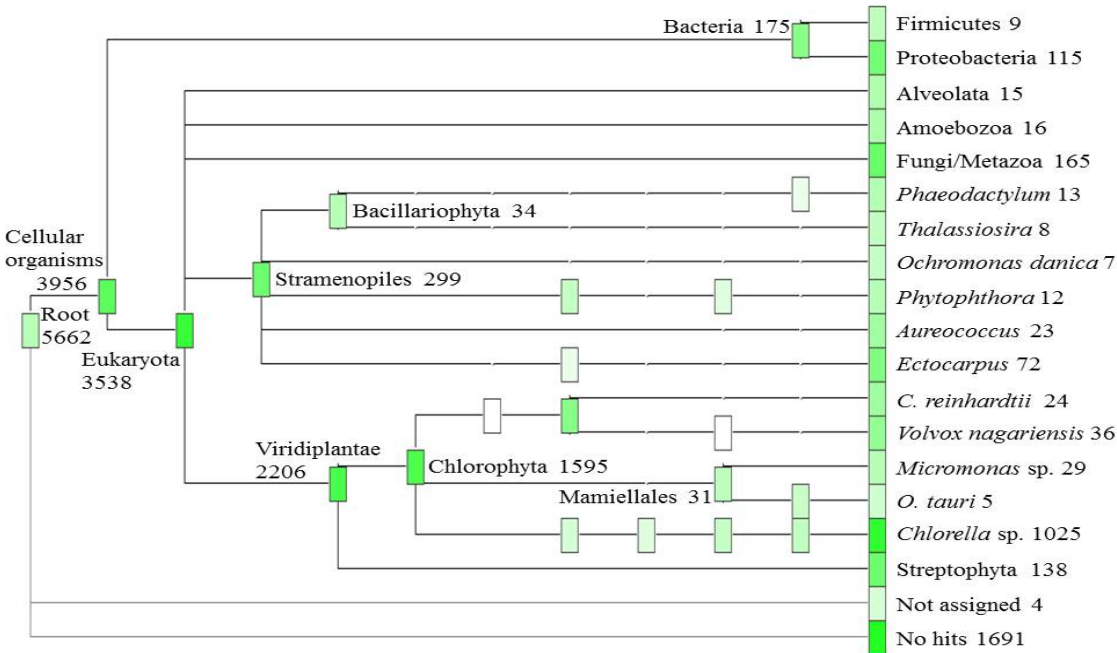
Contig	Phylogeny	% identity
23	<i>Chlorella variabilis</i>	92
285	<i>Chaetosphaeridium globosum</i>	86
577	<i>Chlorella vulgaris</i>	93
737	<i>Chlorella vulgaris</i>	92
1429	<i>Chrysolepidomonas dendrolepidota</i>	97
2150	<i>Mallomonas caudata</i>	89
2177	<i>Chlorella vulgaris</i>	97
2257	<i>Chrysolepidomonas dendrolepidota</i>	87
2539	<i>Plasmodiophora brassicae</i>	86
3079	<i>Poteriochroomonas malhamensis</i>	96
3272	<i>Plasmodiophora brassicae</i>	92
3380	<i>Chlorella variabilis</i>	95
3443	<i>Mallomonas mangofera</i>	97
3745	<i>Plasmodiophora brassicae</i>	89
3774	<i>Chlorella variabilis</i>	98
4875	<i>Chlorella vulgaris</i>	88
5252	<i>Micractinium reisseri</i>	93
5269	<i>Micractinium reisseri</i>	93
5354	<i>Heterosigma akashiwo</i>	90
5394	<i>Sorastrum spinulosum</i>	90
5464	<i>Mallomonas mangofera</i>	94
5541	<i>Chlorella variabilis</i>	94
5611	<i>Phaeobotrys solitaria</i>	96



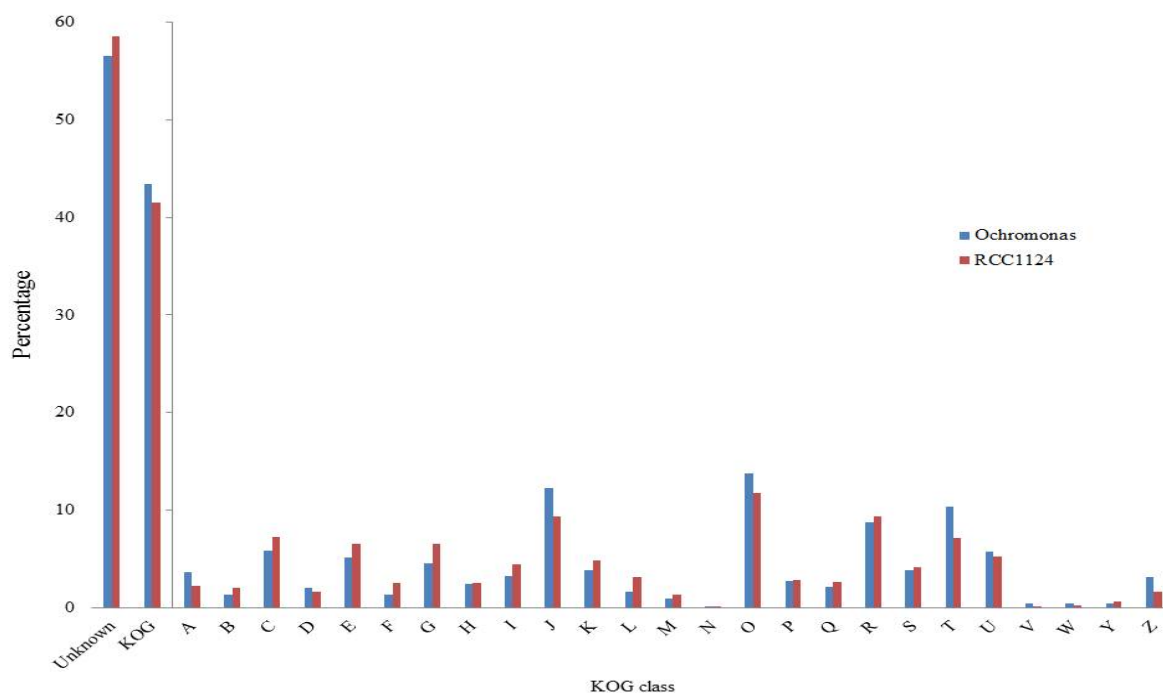
**Figure 6.5. Phylogenetic tree indicating the position of the 18S rRNA contig from the clade VIIA prasinophyte transcriptome (red square). Tree was created using the neighbour joining method with bootstrap analysis (10000 replications)**



**Figure 6.6. Phylogenetic analysis of the top BLASTx hits from the VIIA prasinophyte RCC1124 cDNA-derived contigs with the number of hits assigned to each taxonomic level are indicated.**



**Figure 6.7. Phylogenetic analysis of the top BLASTx hits from the marine *Ochromonas* sp. strain CCMP584 cDNA-derived contigs with the number of hits assigned to each taxonomic level are indicated.**



**Figure 6.8:** The percentage hits against the KOG database and the contribution of KOG classes in the clade VIIA prasinophyte (strain RCC1124) and *Ochromonas* sp. (strain CCMP584) samples. Classes are: A- RNA processing and modification; B - Chromatin structure and dynamics; C - Energy production and conversion; D - Cell cycle control, cell division, chromosome partitioning; E - Amino acid transport and metabolism; F - Nucleotide transport and metabolism; G - Carbohydrate transport and metabolism; H - Coenzyme transport and metabolism; I - Lipid transport and metabolism; J - Translation, ribosomal structure and biogenesis; K - Transcription; L - Replication, recombination and repair; M - Cell wall/membrane/envelope biogenesis; N - Cell motility and secretion O - Posttranslational modification, protein turnover, chaperones; P - Inorganic ion transport and metabolism; Q - Secondary metabolites biosynthesis, transport and catabolism; R - General function prediction only; S - Function unknown; T - Signal transduction mechanisms; U - Intracellular trafficking, secretion, and vesicular transport; V - Defence mechanisms; W - Extracellular structures; Y - Nuclear structure; Z - Cytoskeleton.



### 6.3.6 Light harvesting and carbon fixation

KEGG analysis was undertaken in order to identify the metabolic pathways which were being expressed in each sample. In the prasinophyte sample the majority of the enzymes required for the C<sub>4</sub> dicarboxylic acid cycle were expressed except for a chloroplast targeted malate dehydrogenase NADP<sup>+</sup> (EC:1.1.1.82) and phosphoenolpyruvate carboxykinase (EC:4.1.1.49) (Table 6.3). In the *Ochromonas* cDNAs several of the enzymes required for carbon fixation were not detected in the transcriptome including malate dehydrogenase NADP<sup>+</sup> (EC:1.1.1.82), phosphoenolpyruvate carboxylase (EC:4.1.1.31) and phosphoenolpyruvate carboxykinase (EC:4.1.1.49) (Table 6.4).

In the prasinophyte sample 2 contigs were observed which were identified as transcripts encoding the small subunit of RuBisCO. One of the transcripts was related to *Chlamydomonas* with the other being related to *Chlorella* (Trebouxiophyceae) (Table 6.3).

In the clade VIIA prasinophyte sample two contigs related to  $\alpha$ -carbonic anhydrases were observed. One was phylogenetically related to the green alga *Chlamydomonas* while the other was related to the cyanobacterium *Acetabularia*. The *Ochromonas* cDNAs also contained two  $\alpha$ -carbonic anhydrases related to the bacteria *Mesorhizobium* and *Vibrio*. The *Ochromonas* transcriptome contained 3  $\gamma$  carbonic anhydrases two of which had a top blast hit related to the plant *Arabidopsis* while the third had a hit against the brown alga *Ectocarpus* (Table 6.5).

**Table 6.3. Transcripts involved in carbon fixation in the clade VIIA prasinophyte.**

Enzyme	Contig	Top hit	E value	% positives	Length
aspartate aminotransferase	Contig3344	<i>Volvox</i>	2E-41	79	116
	Contig464	<i>Ricinus</i>	1E-30	59	209
	Contig1562	<i>Chlamydomonas</i>	2E-72	65	211
	Contig465	<i>Ricinus</i>	9E-16	65	113
alanine transaminase	Contig1821	<i>Micromonas</i>	7E-100	79	223
	Contig3101	<i>Capsicum</i>	9E-45	69	156
fructose-bisphosphate aldolase, class I	Contig1208	<i>Ostreococcus</i>	3E-68	86	166
	Contig1725	<i>Saccharomonospora</i>	3E-68	66	342
	Contig1300	<i>Chlamydomonas</i>	1E-15	85	355
	Contig2682	<i>Micromonas</i>	2E-29	74	76
	Contig4008	<i>Ostreococcus</i>	2E-27	81	77
glyceraldehyde-3-phosphate dehydrogenase	Contig2034	<i>Cladophora</i>	2E-115	79	159
	Contig7	<i>Botryococcus</i>	0E+00	88	352
malate dehydrogenase	Contig1122	<i>Micromonas</i>	3E-33	81	137
	Contig2219	<i>Oryza</i>	1E-71	83	159
	Contig2139	<i>Thalassiosira</i>	2E-88	73	226
	Contig4735	<i>Chlamydomonas</i>	2E-15	61	90
	Contig818	<i>Ostreococcus</i>	1E-51	75	169
	Contig820	<i>Chlamydomonas</i>	1E-30	70	115
phosphoglycerate kinase	Contig1964	<i>Synechococcus</i>	1E-113	91	257
phosphoenolpyruvate carboxylase	Contig2194	<i>Micromonas</i>	1E-13	71	92
Phosphoribulokinase	Contig3589	<i>Chlorella</i>	1E-98	94	206
	Contig489	<i>Chlorella</i>	3E-42	81	74
pyruvate kinase	Contig2958	<i>Micromonas</i>	1E-68	84	186
	Contig2981	<i>Ricinus</i>	3E-12	72	65
	Contig3300	<i>Chlamydomonas</i>	1E-69	72	130
	Contig412	<i>Nicotina</i>	7E-08	65	65
	Contig413	<i>Camellia</i>	2E-13	71	77
	Contig414	<i>Oryza</i>	1E-82	78	202
	Contig506	<i>Arabidopsis</i>	5E-26	76	85
pyruvate, orthophosphate dikinase	Contig2149	<i>Ostreococcus</i>	4E-76	78	161
ribose 5-phosphate isomerase A	Contig2330	<i>Thalassiosira</i>	2E-26	79	61
ribulose-bisphosphate carboxylase/oxygenase	Contig1280	<i>Chlamydomonas</i>	2E-142	81	369
	Contig1037	<i>Chlorella</i>	5E-04	72	367
sedoheptulose-bisphosphatase	Contig6	<i>Volvox</i>	1E-158	89	311
ribulose-phosphate 3-epimerase	Contig4633	<i>Prochlorococcus</i>	1E-140	87	243
Transketolase	Contig1100	<i>Volvox</i>	0E+00	75	555

**Table 6.4. Transcripts involved in carbon fixation in the *Ochromonas* sp. CCMP584 transcriptome.**

Enzyme	Contig	Top hit	Evalue	% positives	Length of hit
aspartate aminotransferase	3137	<i>Volvox</i>	3E-62	81	176
fructose-1,6-bisphosphatase I	1233	<i>Solanum</i>	1E-72	76	224
	3969	<i>Volvox</i>	1E-66	89	148
	4443	<i>Karenia</i>	2E-33	80	100
fructose-bisphosphate aldolase, class I	362	<i>Chlamydomonas</i>	2E-109	88	250
	718	<i>Arabidopsis</i>	1E-65	74	225
	1405	<i>Arabidopsis</i>	4E-53	71	130
	3482	<i>Galdieria</i>	2E-38	77	133
glyceraldehyde-3-phosphate dehydrogenase	2711	<i>Mallomonas</i>	3E-57	93	145
	328	<i>Solanum</i>	2E-92	88	246
	1086	<i>Achlya</i>	9E-71	85	204
	2701	<i>Mallomonas</i>	5E-74	87	175
	388	<i>Chlorella</i>	1E-111	94	227
	3706	<i>Coprinopsis</i>	3E-52	78	152
malate dehydrogenase	311	<i>Cicer</i>	4E-21	77	87
	463	<i>Volvox</i>	6E-88	80	265
	1756	<i>Chlorella</i>	2E-92	89	222
	2731	<i>Phytophthora</i>	8E-61	80	175
	4457	<i>Zea</i>	2E-47	89	131
	5006	<i>Volvox</i>	3E-52	90	125
	5593	<i>Pinus</i>	3E-13	81	54
malate dehydrogenase (oxaloacetate-decarboxylating)	3620	<i>Chlamydomonas</i>	1E-27	58	172
phosphoglycerate kinase	2543	<i>Synechococcus</i>	1E-55	87	141
pyruvate kinase	1627	<i>Volvox</i>	2E-27	77	129
pyruvate, orthophosphate dikinase	55	<i>Phytophthora</i>	6E-64	78	206
	1772	<i>Finegoldia</i>	7E-40	76	140
	2752	<i>Volvox</i>	2E-58	73	192
	3906	<i>Ostreococcus</i>	4E-35	71	133
	5028	<i>Congregibacter</i>	9E-16	59	116
	5176	<i>Congregibacter</i>	1E-15	63	98
	5178	<i>Maribacter</i>	1E-25	78	113
	5336	<i>Odoribacter</i>	7E-24	78	89
ribose 5-phosphate isomerase A	840	<i>Chlamydomonas</i>	4E-111	94	234
ribulose-bisphosphate carboxylase/oxygenase	624	<i>Chlamydomonas</i>	3E-92	86	245
	2156	<i>Chlamydomonas</i>	1E-63	86	168
	4168	<i>Chlorella</i>	3E-44	78	134
ribulose-phosphate 3-epimerase	315	<i>Chlamydomonas</i>	8E-100	91	222
Transketolase	53	<i>Chlamydomonas</i>	3E-128	79	351
	714	<i>Salpingoeca</i>	1E-36	58	212
	2849	<i>Lyngbya</i>	2E-43	78	148

**Table 6.5. Transcripts related to carbonic anhydrases in both transcriptomes.**

Contig	Carbonic anhydrase	Phylogeny	Evalue	% positives	Length of hit
491	$\gamma$	<i>Arabidopsis</i>	2E-45	67	219
1250	$\gamma$	<i>Arabidopsis</i>	4E-62	70	222
1715	$\gamma$	<i>Ectocarpus</i>	2E-22	55	175
2326	$\alpha$	<i>Mesorhizobium</i>	5E-07	67	43
3164	$\alpha$	<i>Vibrio</i>	2E-09	54	110
431	$\alpha$	<i>Chlamydomonas</i>	9E-13	56	117
1927	$\alpha$	<i>Acetabularia</i>	8E-14	57	95

The composition of transcripts encoding light harvesting complex proteins also differed between the two algal transcriptomes. Amongst the clade VIIA prasinophyte cDNAs only transcripts encoding chlorophyll a/b binding proteins typical of other Chlorophyta species were observed. Transcripts encoding 6 antenna proteins from photosystem I were observed relating to 5 different types (Table 6.6). Five transcripts encoding antenna proteins of photosystem II were present in the transcriptome (Table 6.6). As well as the antenna proteins a LHC related protein PsbS was also encoded in the transcriptome.

In contrast, the *Ochromonas* sp. CCMP584 transcriptome contained both chlorophyll a/b binding proteins (12 contigs) as well as fucoxanthin chlorophyll a/c binding proteins (4 contigs) (Table 6.7). The fucoxanthin chlorophyll a/c binding proteins were related to *Phaeodactylum* (3 contigs) and a single contig to *Heterosigma*.

**Table 6.6. LHC associated transcripts in the clade VIIA prasinophyte transcriptome**

Contig	Hit	Phylogeny	E value	% positives	Length of hit
526	lhca2	<i>Ricinus</i>	4E-33	63	125
3385	lhca2	<i>Volvox</i>	1E-47	74	151
657	lhca3	<i>Populus</i>	3E-74	71	251
1095	lhca4	<i>Volvox</i>	2E-60	69	193
2578	lhca5	<i>Volvox</i>	6E-56	56	66
2082	lhca6	<i>Chlamydomonas</i>	4E-56	56	196
1235	lhcb	<i>Acetabularia</i>	9E-55	80	115
299	lhcb2	<i>Bigeloviella</i>	3E-103	79	263
2127	lhcb2	<i>Bigeloviella</i>	1E-43	86	112
1253	lhcb4	<i>Chlamydomonas</i>	1E-68	72	230
146	lhcb5	<i>Chlamydomonas</i>	4E-90	83	255
4498	psbS	<i>Acetabularia</i>	7E-78	79	197

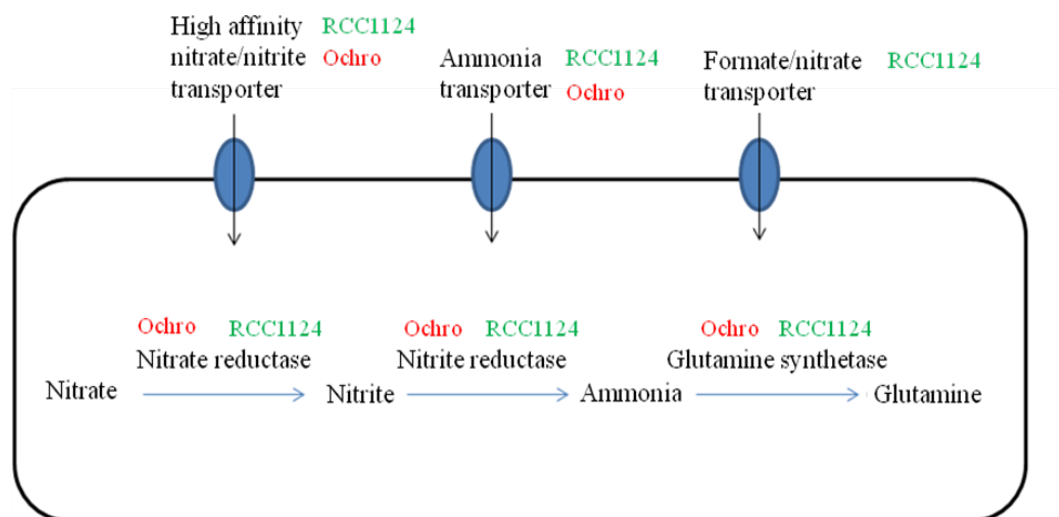
**Table 6.7. Transcripts encoding light harvesting complex proteins in the *Ochromonas* sp. CCMP584 transcriptome.**

Contig	Hit	Phylogeny	E-value	% positives	Length of hit
77	chlorophyll a-b binding protein	<i>Tetraselmis</i>	2E-88	82	256
116	chlorophyll a-b binding protein	<i>Beta</i>	3E-86	82	218
266	chlorophyll a-b binding protein	<i>Chlamydomonas</i>	9E-75	74	235
290	chlorophyll a-b binding protein	<i>Bigeloviella</i>	3E-74	74	227
344	chlorophyll a-b binding protein	<i>Scenedesmus</i>	1E-93	85	224
356	chlorophyll a-b binding protein	<i>Chlamydomonas</i>	3E-100	81	255
732	chlorophyll a-b binding protein	<i>Chlamydomonas</i>	1E-86	85	227
1495	chlorophyll a-b binding protein	<i>Chlamydomonas</i>	2E-64	75	193
2377	chlorophyll a-b binding protein	<i>Volvox</i>	2E-59	75	179
2439	chlorophyll a-b binding protein	<i>Chlamydomonas</i>	9E-69	80	175
3047	chlorophyll a-b binding protein	<i>Volvox</i>	3E-56	78	152
4314	chlorophyll a-b binding protein	<i>Arabidopsis</i>	2E-37	76	122
1939	fucoxanthin chlorophyll a/c protein	<i>Phaeodactylum</i>	2E-32	67	131
2867	fucoxanthin chlorophyll a/c protein	<i>Phaeodactylum</i>	1E-15	70	89
5065	fucoxanthin chlorophyll a/c protein	<i>Phaeodactylum</i>	1E-15	66	87
5202	fucoxanthin chlorophyll a/c protein	<i>Heterosigma</i>	8E-14	88	49

### 6.3.7 Nutrient acquisition and assimilation

#### 6.3.7.1 Nitrogen metabolism

Both the clade VII prasinophyte and *Ochromonas* transcriptomes contained cDNAs for various aspects of nitrogen metabolism (see Figure 6.9). This included both nitrate and nitrite reductases. Specifically, in the prasinophyte transcriptome 5 nitrogen transporters were present including a low/dual affinity formate/nitrate transporter, 2 high affinity nitrate/nitrite transporters (MFS) and 2 ammonium transporters (Table 6.8). Transcripts encoding nitrate reductase, nitrite reductase and glutamine synthase were all found in the *Ochromonas* transcriptome. In the *Ochromonas* transcriptome 4 transporters were observed with 2 ammonium transporters and two nitrate transporters (Table 6.9).



**Figure 6.9.** Encoded enzymes and transporters involved in the acquisition and assimilation of nitrogen within either the clade VIIA prasinophyte (labelled RCC1124) or *Ochromonas* sp. CCMP584 (labelled Ochro) transcriptome.

**Table 6.8. Transcripts encoding genes involved in nitrogen assimilation and utilisation in the clade VIIA prasinophyte transcriptome**

Contig	Hit	Phylogeny	E value	% positives	Length of hit
Contig1669	aminomethyltransferase	<i>Arabidopsis</i>	9E-42	81	129
Contig3556	aminomethyltransferase	<i>Flaveria</i>	2E-41	75	137
Contig3320	ammonium transporter	<i>Micromonas</i>	6E-12	63	91
Contig2436	ammonium transporter	<i>Micromonas</i>	3E-12	60	111
Contig1087	asparagine synthase (glutamine-hydrolysing)	<i>Ostreococcus</i>	1E-43	77	131
Contig392	asparagine synthase (glutamine-hydrolysing)	<i>Helianthus</i>	1E-31	69	150
Contig897	asparagine synthase (glutamine-hydrolysing)	<i>Micromonas</i>	4E-71	77	225
Contig1726	carbamoyl-phosphate synthase (glutamine)	<i>Synechococcus</i>	5E-47	76	123
Contig868	cystathionine gamma-lyase	<i>Ostreococcus</i>	8E-145	80	387
Contig821	ferredoxin-nitrite reductase	<i>Nicotiana</i>	3E-173	73	531
Contig3773	formate-dependent nitrate reductase	<i>Escherichia</i>	4E-49	100	60
Contig2282	formate-dependent nitrite reductase, Fe-S protein	<i>Shigella</i>	3E-44	100	82
Contig321	formate-nitrite transporter	<i>Ostreococcus</i>	7E-55	75	252
Contig4600	glutamate dehydrogenase (NAD(P)+)	<i>Chlamydomonas</i>	1E-174	76	424
Contig1530	glutamate dehydrogenase (NADP+)	<i>Psychroflexus</i>	3E-39	66	173
Contig2631	glutamate dehydrogenase (NADP+)	<i>Leadbetterella</i>	3E-22	58	145
Contig4070	glutamate dehydrogenase (NADP+)	<i>Prevotella</i>	8E-43	77	118
Contig2204	glutamate synthase (ferredoxin)	<i>Micromonas</i>	7E-101	77	286
Contig1977	glutamate synthase (NADPH/NADH) large chain	<i>Shigella</i>	2E-62	90	144
Contig1353	glutaminase	<i>Capsaspora</i>	1E-10	58	125
Contig2484	glutamine synthetase	<i>Saccharum</i>	9E-128	80	223
Contig925	glutamine synthetase	<i>Chlamydomonas</i>	4E-18	88	60
Contig1999	high affinity nitrate transporter (MFS)	<i>Micromonas</i>	3E-90	72	301
Contig3061	L-asparaginase	<i>Micromonas</i>	6E-38	72	130
Contig1960	MFS family transporter: nitrate/nitrite	<i>Zea</i>	4E-24	85	39
Contig1261	nitrate reductase (NADH)	<i>Medicago</i>	0E+00	70	694
Contig2381	nitrate reductase (NADH)	<i>Streptomyces</i>	2E-24	68	115
Contig1155	nitrogenase (flavodoxin)	<i>Prochlorococcus</i>	2E-47	72	168

**Table 6.9. Transcripts encoding genes involved in nitrogen assimilation and utilisation in the *Ochromonas* sp. CCMP584 transcriptome**

Hit	Phylogeny	E-value	% positives	Length of hit
ABC-type nitrate transporter	<i>Marinobacter</i>	2E-71	99	134
ammonium transporter	<i>Micromonas</i>	9E-51	61	264
ammonium transporter	<i>Alkaliphilus</i>	1E-07	61	59
asparagine synthase (glutamine-hydrolysing)	<i>Micromonas</i>	3E-53	72	183
carbamoyl-phosphate synthase (glutamine)	<i>Chlamydomonas</i>	1E-61	78	211
ferredoxin-nitrite reductase	<i>Chlamydomonas</i>	4E-66	78	206
formamidase	<i>Paenibacillus</i>	4E-37	68	152
glutamate dehydrogenase	<i>Ectocarpus</i>	1E-36	64	207
glutamate synthase (ferredoxin)	<i>Chlamydomonas</i>	3E-104	80	274
glutamate synthase (NADPH)	<i>Arabidopsis</i>	4E-67	80	181
glutamate synthase (NADPH/NADH)	<i>Ectocarpus</i>	1E-45	77	137
glutamate synthase (NADPH/NADH)	<i>Arabidopsis</i>	6E-51	86	134
glutamine synthetase	<i>Prototheca</i>	3E-61	92	131
high affinity nitrate transporter	<i>Chlorella</i>	5E-55	76	181
nitrate reductase (NADH)	<i>Chlorella</i>	3E-80	77	247
nitrate reductase (NADH)	<i>Marinobacter</i>	2E-73	82	206

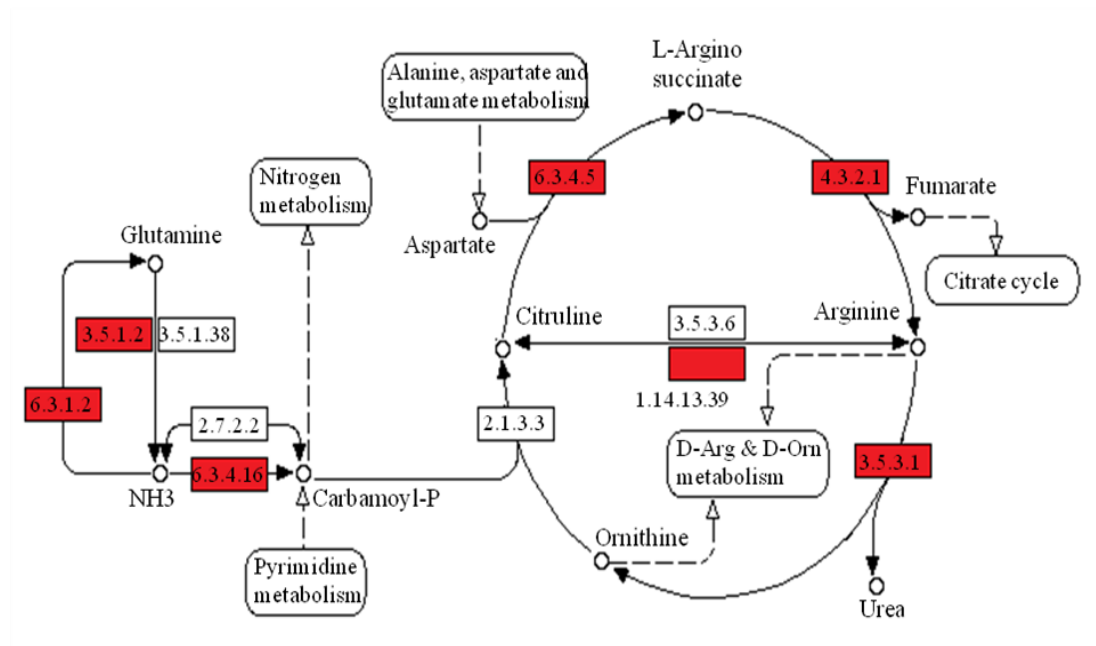
Within the *Ochromonas* transcriptome transcripts encoding the majority of the enzymes involved in the urea cycle were predicted against the KEGG database (Figure 6.10 and Table 6.10). The only enzyme for which transcripts were not found was ornithine carbamoyl transferase which converts ornithine to citrulline. Transcripts encoding the two enzymes, glutamine synthase and glutaminase were present. These encode the reversible reaction between glutamine and glutamate.



**Table 6.10. Transcripts in the *Ochromonas* sp. CCMP584 transcriptome encoding components of the urea cycle when analysed against the KEGG database.**

Contig	KEGG hit	KEGG ID	Top KEGG hit	E value
contig03711	nitric-oxide synthase	EC:1.14.13.39	<i>Drosophila</i>	9E-09
contig00444	nitric-oxide synthase	EC:1.14.13.39	<i>Nasonia</i>	7E-09
contig00653	Glutaminase	EC:3.5.1.2	<i>Branchiostoma</i>	5E-06
contig01666	Glutaminase	EC:3.5.1.2	<i>Monosiga</i>	8E-07
contig03135	Arginase	EC:3.5.3.1	<i>Fusobacterium</i>	1E-05
contig05030	argininosuccinate lyase	EC:4.3.2.1	<i>Ostreococcus</i>	3E-32
contig00484	argininosuccinate lyase	EC:4.3.2.1	<i>Saccharopolyspora</i>	4E-15
contig05274	argininosuccinate lyase	EC:4.3.2.1	<i>Chlamydomonas</i>	2E-24
contig01566	glutamine synthetase	EC:6.3.1.2	<i>Chlamydomonas</i>	1E-50
contig00246	glutamine synthetase	EC:6.3.1.2	<i>Clostridium</i>	7E-47
contig01402	carbamoyl-phosphate synthase	EC:6.3.4.16	<i>Xenopus</i>	2E-09
contig02946	argininosuccinate synthase	EC:6.3.4.5	<i>Chlamydomonas</i>	2E-80

A contig attributable to carbamoyl phosphate synthase was also present in the *Ochromonas* transcriptome, though it was phylogenetically affiliated with the Chlorophyta against the NCBI nr database. The carbamoyl phosphate synthase present in the transcriptome was a glutamine dependent type as suggested by analysis against the NCBI conserved domain database. Transcripts encoding the enzymes argininosuccinate synthase, argininosuccinate lyase and arginase were also detected (Figure 6.10).



**Figure 6.10.** Transcripts encoding enzymes of the urea cycle in *Ochromonas* sp. CCMP584 (EC numbers of enzymes are highlighted in red). EC numbers: 6.3.1.2 = glutamine synthetase, 3.5.1.2 = glutaminase, 3.5.1.38 = glutamin-(asparagin-)-ase, 2.7.2.2 = carbamate kinase, 6.3.4.16 = carbamoyl-phosphate synthase (ammonia), 2.1.3.3 = ornithine carbamoyltransferase, 6.3.4.5 = argininosuccinate synthase, 3.5.3.6 = arginine deiminase, 1.14.13.39 = nitric-oxide synthase, 3.5.3.1 = arginase, 4.3.2.1 = argininosuccinate lyase.

#### 6.3.7.2 Sulphur metabolism

Both the clade VIIA prasinophyte and *Ochromonas* sp. transcriptomes contained transcripts encoding enzymes involved in sulphur metabolism. Both organisms appear to assimilate sulphate to sulphite for subsequent cysteine biosynthesis (Table 6.11 and 6.12). Both transcriptomes also include secondary sulphate transporters (SulP) whilst the clade VIIA prasinophyte transcriptome also encodes an ATP-dependent Fe/Sulphur ABC transporter (Table 6.11). The *Ochromonas* sample has a dual sodium/sulphate symporter (DASS) in addition to the SulP transporter (Table 6.12).

**Table 6.11. Assimilation and transport of sulphur in the clade VIIA prasinophyte transcriptome.**

Contig	Hit	Phylogeny	E value	% positives	Length of hit
Contig699	sulphite reductase (NADPH) flavoprotein alpha-component	<i>Ostreococcus</i>	2E-90	81	233
Contig1155	sulphite reductase (NADPH) flavoprotein alpha-component	<i>Prochlorococcus</i>	2E-47	72	168
Contig1752	sulphate adenylyltransferase	<i>Ostreococcus</i>	4E-87	87	163
Contig1753	sulphate adenylyltransferase	<i>Chlamydomonas</i>	2E-97	88	214
Contig2109	cysteine synthase A	<i>Ostreococcus</i>	4E-23	88	101
Contig2381	sulphite oxidase	<i>Arthrobacter</i>	8E-25	71	112
Contig4321	sulphate adenylyltransferase	<i>Ostreococcus</i>	7E-22	94	55
Contig4407	sulphite reductase (NADPH) flavoprotein	<i>Chlamydomonas</i>	3E-18	79	61
Contig1644	Sulphate transporter (SulP)	<i>Populus</i>	2E-86	66	157
Contig1844	ABC SufBCD system	<i>Micromonas</i>	0E+00	92	389

**Table 6.12. Assimilation and transport of sulphur in the *Ochromonas* sp. CCMP584 transcriptome.**

Contig	Hit	Phylogeny	E-value	% positives	Length of hit
contig00056	cysteine synthase	<i>Ostreococcus</i>	2E-111	83	314
contig01146	sulfite reductase (ferredoxin)	<i>Microcoleus</i>	5E-52	68	190
contig02187	sulfite oxidase	<i>Camponotus</i>	2E-08	63	78
contig03711	sulfite reductase	<i>Ectocarpus</i>	8E-70	87	168
contig05036	cysteine synthase	<i>Polytomella</i>	5E-27	87	92
contig00663	cysteine synthase	<i>Chlamydomonas</i>	1E-33	84	79
contig00466	high affinity sulfate transporter	<i>Chlamydomonas</i>	2E-68	71	284
contig01762	Sulfate transporter (SulP)	<i>Ricinus</i>	3E-44	72	162
contig03577	sodium ion/sulfate	<i>Ectocarpus</i>	2E-37	72	149

### 6.3.7.3 Cellular iron utilisation

In the clade VIIA prasinophyte transcriptome only transcripts encoding manganese superoxide dismutases (SODs) were detected (Table 6.13), whilst in the *Ochromonas* transcriptome both iron and manganese SODs were detectable (Table 6.14).

**Table 6.13. Iron utilisation (and cellular alternatives) in the clade VIIA prasinophyte transcriptome**

Contig	Hit	Phylogeny	E-value	% positives	Length of hit
579	Mn superoxide dismutase	<i>Microcoleus</i>	4E-08	56	55
585	Mn superoxide dismutase	<i>Micromonas</i>	1E-62	65	120
1843	Plastocyanin	<i>Scenedesmus</i>	3E-41	83	134
298	Ferredoxin	<i>Cyanothece</i>	8E-17	64	87
1309	Ferredoxin	<i>Oryza</i>	5E-31	86	99
1431	Ferredoxin	<i>Chlamydomonas</i>	6E-20	66	121
1155	Flavodoxin	<i>Prochlorococcus</i>	2E-47	72	168

Transcripts encoding plastocyanin were the only type of electron carrier between the cytochrome  $b_6/f$  complex and photosystem I that were detectable in the clade VIIA transcriptome, with transcripts encoding the iron requiring cytochrome  $c_6$  being undetectable in the transcriptome.

**Table 6.14. Iron utilisation (and cellular alternatives) in the *Ochromonas* sp. CCMP584 transcriptome.**

Contig	Hit	Phylogeny	E-value	% positives	Length of hit
1084	Mn superoxide dismutase	<i>Haematococcus</i>	1E-68	75	215
1379	Mn superoxide dismutase	<i>Synechococcus</i>	1E-64	74	197
2385	Fe superoxide dismutase	<i>Chlamydomonas</i>	6E-58	80	153
2898	Plastocyanin	<i>Pediastrum</i>	1E-44	83	122
810	Ferredoxin	<i>Volvox</i>	3E-68	75	236
1439	Ferredoxin	<i>Ectocarpus</i>	1E-39	80	158
2941	Ferredoxin	<i>Chlamydomonas</i>	2E-32	71	120
3649	Ferredoxin	<i>Ectocarpus</i>	3E-39	77	125
3779	Ferredoxin	<i>Ectocarpus</i>	3E-07	49	111
4210	Ferredoxin	<i>Chlamydomonas</i>	1E-28	74	109
3381	Flavodoxin	<i>Chlamydomonas</i>	1E-50	70	158

A similar observation was noted for the *Ochromonas* sample where cytochrome  $c_6$  derived transcripts were also not detectable. In both the clade VIIA prasinophyte and *Ochromonas* transcriptomes contigs affiliated to the

electron transfer proteins ferredoxin and flavodoxin were identified. The *Ochromonas* sample also contained transcripts encoding components of a high affinity iron transport system including ferric reductase and a ferric permease although no multicopper ferroxidase was detectable. In the clade VIIA prasinophyte transcriptome a ferric reductase was detectable, along with a putative siderophore synthesis protein related to an enterobactin from *Brevibacterium*. Also present in the clade VIIA prasinophyte transcriptome was an iron specific ABC transporter related to *Ostreococcus* (Table 6.15).

**Table 6.15. Transcripts involved in the acquisition of iron in the two transcriptomes.**

Contig	Hit	Phylogeny	E-value	% positives	Length of hit
642	Ferric reductase	<i>Phaeodactylum</i>	4E-09	60	99
5117	Ferric permease	<i>Citrobacter</i>	5E-55	100	115
3276	Ferric reductase	<i>Shigella</i>	3E-35	100	75
1567	Enterobactin	<i>Brevibacterium</i>	6E-05	64	58
1290	Fe ABC transporter	<i>Ostreococcus</i>	1E-66	75	219

## **6.4 Discussion**

Work performed in this chapter set out to provide an indication of the functional potential of two environmentally relevant PPEs, as evidenced from construction of cDNA libraries, an approach that is beginning to be performed for several algal taxa (e.g. see Mock et al., 2006; Phillips et al., 2008; Shi et al., 2008; Terauchi et al., 2010). It was intended that the availability of ‘expressed sequence tag’ libraries for a clade VIIA prasinophyte and a marine chrysophyte would also aid in the assignment of

environmental metatranscriptome reads (see chapter 7). Further discussion of some of the more interesting findings obtained from these transcriptome libraries is now presented.

#### **6.4.1 Transcriptome statistics and phylogenetics**

The current study aimed to investigate the transcriptome of two environmentally relevant PPEs. The selected targeting of the poly(A) tail of eukaryotic mRNA allowed for the increased analysis of functional traits expressed by each organism. In order to do this poly (dT) magnetic beads were used to enrich for mRNAs in the sample compared to rRNAs. As a result, contigs attributable to rRNAs accounted for only 0.3% and 1.2% of total transcripts for the prasinophyte and *Ochromonas* samples, respectively. Other studies have also shown the benefit of specific selection for mRNAs e.g. a study investigating gene expression in algal mats (Grant *et al.*, 2006) showed rRNA accounted for 60% of the transcriptome when there was no selection, but that this value fell to only 4% in ‘selected’ samples. The current study shows even lower ‘contamination’ values for rRNA suggesting that the use of magnetic beads for purification of mRNA will be widely applicable for the transcriptomic analysis of unicellular algal cultures.

Phylogenetic analysis of the rRNA contigs present in the clade VIIA prasinophyte sample (see section 6.3.2) showed their affiliation (98% identity) to a 18S rRNA sequence from another clade VIIA prasinophyte strain, RCC287, isolated from the Pacific Ocean (see <http://www.sb-roscoff.fr/Phyto/RCC/>). However, formal description of these organisms as

novel genera/species is still awaited. In contrast, for the *Ochromonas* sample rRNA contigs were affiliated with *Chlorella* (Trebouxiophyceae) and a putative cercozoan heterotroph, as well as hits to itself i.e. *Ochromonas* sp. CCMP584. This suggests that the latter culture contained other contaminating species despite the fact that prior to transcriptome analysis, sequencing of 18S rRNA PCR products indicated a uni-algal culture (see section 6.3.1). The contamination in the *Ochromonas* culture makes it difficult to assign metabolic pathways to specific taxon.

Phylogenetic analysis of the clade VIIA prasinophyte transcriptome indicated that the majority of the top BLASTx hits were against Viridiplantae and especially the Chlorophyta lineage which includes the Prasinophyceae. However, a high percentage of the reads could not be assigned to a taxon. For the *Ochromonas* transcriptome the majority of the assigned reads were related to *Chlorella* (Trebouxiophyceae) but with a significant proportion of the hits (~8%) related to stramenopiles. However, only a small proportion of the contigs were actually related to *Ochromonas*. As with the prasinophyte sample a large number of the reads were unable to be assigned to a specific taxon. This could be due to the lack of genetic information present in databases related to chrysophyte species.

Neither culture was axenic with the *Ochromonas* culture especially having a high bacterial load (data not shown). The *Ochromonas* strain was cultured using a grain of rice to promote the growth of heterotrophic bacteria as a food source for this mixotrophic organism. Despite the high bacterial load of the culture, the low percentage of transcript reads attributable to bacteria

in both samples suggests again that the selective targeting of eukaryotic mRNAs using poly(dT) magnetic beads was very efficient.

The clade VIIA prasinophyte transcriptome showed a higher average %GC content of reads compared to the *Ochromonas* sample (Figure 6.4). However, this higher %GC content of the clade VIIA prasinophyte reads is similar to that observed in another prasinophyte genome *Ostreococcus tauri* (59%) (Derelle *et al.*, 2006). The main peak in %GC content of reads for the *Ochromonas* sample was ~45%. This is lower than the %GC content previously found for the freshwater strain *Ochromonas danica*, 49.9% (Terauchi *et al.*, 2010 supplementary info). However, the %GC content of reads in the *Ochromonas* sample was considerably lower than that known for *Chlorella* at 67% although there was a small peak in the %GC at 65% which could have related to *Chlorella* (Blanc *et al.*, 2010).

#### 6.4.2 Light harvesting and carbon fixation

All marine algae employ carbon concentrating mechanisms (CCM) to concentrate CO<sub>2</sub> at the active site of RuBisCO (Raven *et al.*, 2008). Of these, a CCM system comprising a bicarbonate transporter and a carbonic anhydrase is perhaps the best known. There are 5 known types of carbonic anhydrases: alpha ( $\alpha$ ), beta ( $\beta$ ), gamma ( $\gamma$ ), delta ( $\delta$ ) and epsilon ( $\epsilon$ ) (Parker *et al.*, 2008). Only the  $\epsilon$ -carbonic anhydrase has not been observed in unicellular algae (Parker *et al.*, 2008). The clade VIIA prasinophyte transcriptome contained reads affiliated with two  $\alpha$ -carbonic anhydrases (section 6.3.6). This is similar to the situation in *Micromonas* (Worden *et*



*al.*, 2009) and in the green alga *Chlamydomonas* (Mitra *et al.*, 2004). In these latter two organisms the  $\alpha$ -carbonic anhydrases are lumenal-targeted (as opposed to stromal-targeted) and are known to be essential for growth under ambient CO<sub>2</sub> conditions (Karlsson *et al.*, 1998). Moreover, at least for *Chlamydomonas* the lumenal-targeted CA is necessary for proper CCM function at low inorganic C concentrations by providing an ample supply of CO<sub>2</sub> for RuBisCO. In contrast the single cytosolic  $\beta$ -carbonic anhydrase present in *Ostreococcus* (Derelle *et al.*, 2006) has been proposed to capture leaking CO<sub>2</sub> for re-import of bicarbonate into the plastid (Worden *et al.*, 2009). Thus, it is possible that the properties of the clade VIIA prasinophyte  $\alpha$ -carbonic anhydrase are similar to those found in *Micromonas* and *Chlamydomonas* rather than *Ostreococcus*, but further functional work would be required to verify this.

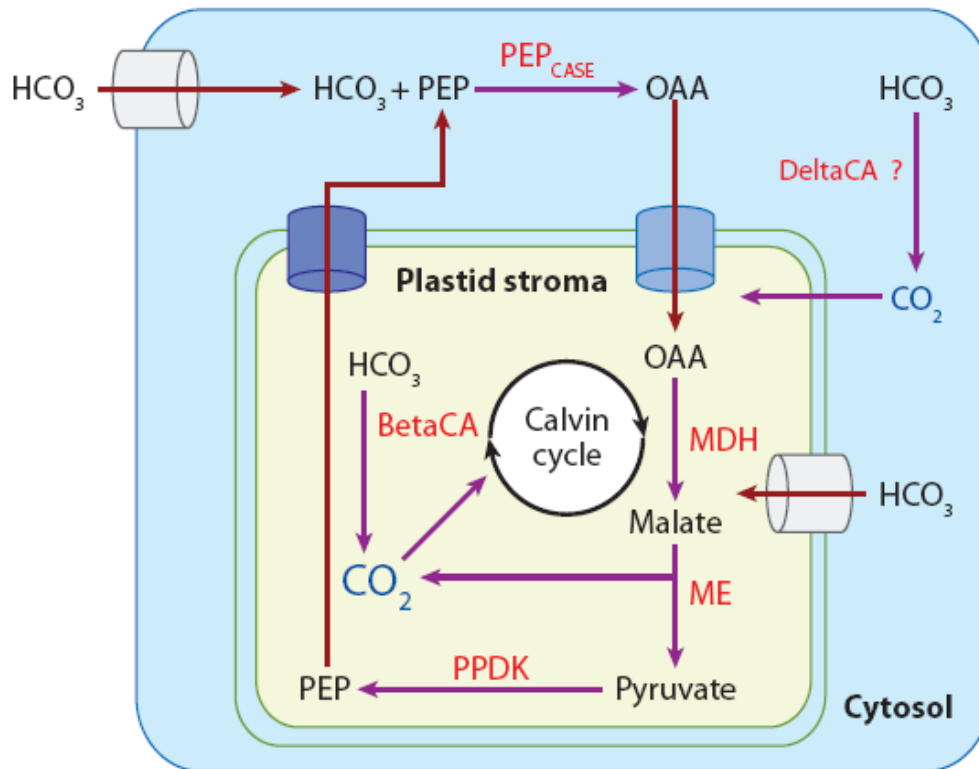
In the *Ochromonas* transcriptome as well as two  $\alpha$ -carbonic anhydrases, three  $\gamma$ -carbonic anhydrases were also present (section 6.3.6). The presence of a  $\gamma$ -carbonic anhydrase is similar to the situation found in the prymnesiophyte *Emiliana huxleyi*, though in this organism it is thought to play a role in calcification rather than to facilitate the concentration of CO<sub>2</sub> near the RuBisCO active site (Soto *et al.*, 2006). Clearly, a diverse range of carbonic anhydrases are observed in marine algae and further investigation of those found to be expressed in the two transcriptomes under study here will be necessary before specific functions can be attributed to them.

With regard to light harvesting, the *Ochromonas* transcriptome contained transcripts encoding light harvesting complex proteins for both chlorophyll

*a/b* and fucoxanthin chlorophyll *a/c* binding proteins (see section 6.3.6). Whilst the fucoxanthin chlorophyll *a/c* binding proteins are typical of those found in chrysophytes (Gibbs and Biggins, 1989), it is presumed that the reads affiliated to chlorophyll *a/b* binding proteins are transcripts arising from the trebouxiphyte contamination. The observation of only chlorophyll *a/b* binding protein reads in the clade VIIA prasinophyte transcriptome is however, typical for green algae. The clade VIIA prasinophyte transcriptome contained a set of 5 different *lhca* genes encoding an LHCI antenna, whilst the *Ostreococcus* genome also contains a small set of 5 *lhca* genes (Derelle *et al.*, 2006). No transcripts were found relating to the the prasinophyte lhcp type antenna proteins which are specific to the Mamiellales (Kozioł *et al.*, 2007). However, a minor PSII antenna protein (LCHB4) was present in the transcriptome which is believed to be ancestral in chlorophyll *a/b* containing organisms (Kozioł *et al.*, 2007). A single transcript encoding PsbS was present in the transcriptome. PsbS is believed to be an important component in dissipation of excess energy via non-photochemical quenching (Bonente *et al.*, 2008).

Interestingly, the clade VIIA prasinophyte transcriptome also contained reads putatively encoding enzymes required for C<sub>4</sub> photosynthesis including phosphoenolpyruvate carboxylase and pyruvate phosphate dikinase. The only exception was that transcripts encoding the chloroplast targeted malate dehydrogenase were not detected. Functional C<sub>4</sub> photosynthesis has yet to be determined unequivocally in unicellular algae, but indications of this capacity have been found in *Ostreococcus* (Derelle *et al.*, 2006),

*Micromonas* (Worden *et al.*, 2009), and in the diatom *Phaeodactylum tricornutum* (Kroth *et al.*, 2008). The C<sub>4</sub> pathway is an alternative mechanism to concentrate CO<sub>2</sub> at the active site of RuBisCO. Different mechanisms have been proposed for C<sub>4</sub> photosynthesis in these organisms. The simplest version proposed for microalgae is found in the prasinophyte *Ostreococcus* (Figure 6.11). The CCM involves the fixation of CO<sub>2</sub> into phosphoenolpyruvate (PEP) and then the generation of oxaloacetate (OAA). OAA is transported into the plastid and converted into malate. This is then decarboxylated by NADP-dependent malic enzymes to generate CO<sub>2</sub> at the active site of RuBisCO. The by-product, pyruvate, is regenerated to PEP to complete the cycle (Parker *et al.*, 2008). A similar mechanism may be utilised by the clade VIIA prasinophyte analysed in this study, although further investigations would be required to determine the localisation of each enzyme. Despite the costs of a C<sub>4</sub> pathway, the presence of this pathway in prasinophytes may compensate for the relatively low CO<sub>2</sub> affinity of their RuBisCO (Parker *et al.*, 2008). Alongside the potential use of an  $\alpha$ -carbonic anhydrase a C<sub>4</sub>-type CCM may thus give these cells a selective advantage in low CO<sub>2</sub> conditions over competitors with less efficient CCMs.



**Figure 6.11. Proposed model for C<sub>4</sub> carbon fixation in *Ostreococcus*.** Detailed explanation in text (Parker *et al.*, 2008). PEPcase: phosphoenolpyruvate carboxylase; PPDK: pyruvate-phosphate dikinase; CA: carbonic anhydrase; MDH: malate dehydrogenase; ME: malic enzyme; OAA: oxaloacetate; PEP: phosphoenolpyruvate

#### 6.4.4 Nutrient Acquisition and Assimilation

Regarding nitrogen acquisition and metabolism, the clade VIIA prasinophyte and *Ochromonas* transcriptomes contained an array of transcripts encoding genes for the metabolism of nitrogen. Like other marine algae (e.g. *Ostreococcus* (see Derelle *et al.*, 2006)), the observed data (see section 6.3.7.1) suggests that these organisms can grow on both oxidised and reduced forms of inorganic nitrogen. In the open ocean ammonium drives the production of phytoplankton (Parker *et al.*, 2008).

Hence, it was not too surprising that the clade VIIA prasinophyte transcriptome contained reads affiliated to two (of the four) *Ostreococcus* ammonium transporters (Derelle *et al.*, 2006), a feature that may indicate it is a strong competitor for this resource.

In the *Ochromonas* transcriptome all but one enzyme required for a functional urea cycle were detected (see section 6.3.7.1). The identification of a functioning urea cycle in a phototroph was first discovered with the sequencing of the two diatom genomes (Armbrust *et al.*, 2004; Bowler *et al.*, 2008). Previously, the presence of a urea cycle was thought to be a unique heterotroph pathway for the detoxification of intracellular ammonia (Armbrust *et al.*, 2004). The form of carbamoyl phosphate synthase present in the *Ochromonas* transcriptome was similar to that of green algae utilising glutamine (see section 6.3.7.1). In the diatom genomes this form has been lost and a carbamoyl phosphate synthase which uses  $\text{NH}_4^+$  is present instead. Very recently, the function of the urea cycle in diatoms has been more clearly elucidated with the idea that it serves as a distribution and repackaging hub for inorganic carbon and nitrogen and contributes significantly to the metabolic response of diatoms to episodic nitrogen availability (Allen *et al.*, 2011). Certainly, with the possible presence of a urea cycle in *Ochromonas* the ability to rapidly recover from prolonged nitrogen limitation may not be confined to diatoms.

For sulphur acquisition and metabolism, an element incorporated into e.g. S-containing amino acids, membrane sulfolipids and cell walls (see Takahashi *et al.*, 2011) but which is non-limiting in seawater, as expected the

prasinophyte clade VIIA and *Ochromonas* transcriptomes contained the metabolic potential for acquisition and metabolism of sulphate (see section 6.3.7.2).

In contrast, productivity in surface waters of the open ocean is often limited by low levels of iron (e.g. see Moore *et al.*, 2002). Growth rate or yield limitation means that phytoplankton require mechanisms to either reduce the cellular demands for iron or have more efficient uptake systems. Whilst evidence for reducing the cellular demand for iron is essentially outside the scope of a transcriptome study, evidence was obtained that both the clade VIIA prasinophyte and the *Ochromonas* sp. have the capacity for high affinity uptake of iron (see section 6.3.7.3). This included the presence of putative ferric reductase transcripts, whose protein products can mediate the initial step of iron uptake by reducing Fe(III) to Fe(II) at the cell surface, followed by the coupled oxidation of Fe(II) and transport of Fe(III). These latter steps are best known in yeast, where the Fe(II) at the cell surface binds to a copper (Cu)-containing ferroxidase (FET3p), where it is oxidized to Fe(III). Ferric Fe is then transported to an Fe(III) permease (FTR1p), which exists at the cell surface as a stable complex with FET3p (reviewed in Kosman 2003). A copper-dependent iron assimilation pathway including a multicopper ferroxidase and a ferric permease has also been suggested in the green alga *Chlamydomonas* (La Fontaine *et al.*, 2002) and in diatoms (Armbrust *et al.*, 2004; Kustka *et al.*, 2007). However, genes encoding components of the high affinity Fe acquisition system appear lacking in *Ostreococcus* genomes (Palenik *et al.*, 2007). The presence of components

of this system in the *Ochromonas* transcriptome here suggests they also acquire iron via this route. In the clade VIIA prasinophyte transcriptome a further system for iron acquisition may be present since transcripts encoding proteins involved in enterobactin synthesis, a bacterial iron scavenging siderophore were found (see section 6.3.7.3). The ability of algae to utilise siderophores for the transport of iron has been previously hypothesised for both diatoms and *Ostreococcus lucimarinus* (Armbrust *et al.*, 2004; Palenik *et al.*, 2007).

Transcripts encoding plastocyanin were the only electron carriers between the cytochrome  $b_6/f$  complex and photosystem I that were expressed in either sample. This is similar to that found following analysis of the *Ostreococcus* genomes (Palenik *et al.*, 2007). Plastocyanin contains a copper atom in order to shuttle electrons between cytochrome  $b_6/f$  complex and photosystem I. No transcripts affiliated to the iron requiring cytochrome  $c_6$  were found, though this carrier is found in cyanobacteria and other eukaryotic algae including *Chlamydomonas*, where it is replaced by plastocyanin in iron-deplete but copper replete conditions (Howe *et al.*, 2006). However, there is a possibility that cytochrome  $c_6$  could be plastid encoded which would not be detected in the transcriptome due to the lack of poly(A) tails in plastid mRNA.

Superoxide dismutases are used to minimise the damage of reactive oxygen species generated during photosynthesis (Bowler *et al.*, 1992). The *Ochromonas* transcriptome contained both iron and manganese SODs (see section 6.3.7.3), a similar situation to that observed in the diatom

*Thalassiosira pseudonana* (Armbrust *et al.*, 2004). However, the iron SOD was phylogenetically related to *Chlorella*, which may suggest that it derived from this ‘contaminant’ species and not *Ochromonas*. In the clade VIIA prasinophyte transcriptome a single manganese SOD was detected, whilst no FeSOD affiliated transcripts were detectable. If truly absent, the lack of a FeSOD may reflect an adaptation to reduce the iron quotas in this organism. For completeness, Cu/Zn SODs have also been shown to be present in both the prasinophyte *Ostreococcus* and the diatom *P. tricornutum* (Parker *et al.*, 2008), but transcripts for this SOD were not detected in either transcriptome analysed here.

Further reduction in the iron quota of the clade VIIA prasinophyte may be manifest by the presence of transcripts for both ferredoxin and flavodoxin, with the latter photosynthetic electron transporter lacking iron, and known to replace ferredoxin at least in *Ostreococcus* (Palenik *et al.*, 2007).

## **6.5 Conclusions and further work**

Analysis of the Clade VIIA prasinophyte and *Ochromonas* transcriptomes has highlighted some of the metabolic pathways important in carbon fixation and nutrient assimilation in two ecologically important PPEs. The clade VIIA prasinophyte (strain RCC1124) appears to express genes used in the C<sub>4</sub> photosynthesis pathway whilst both organisms contain components of a biophysical CCM utilising carbonic anhydrases, and have the capacity to acquire and assimilate a variety of nitrogen forms. However, due to the



contamination in the *Ochromonas* culture the metabolic pathways present in the culture could not be assigned to a specific taxon.

The current work only analysed the transcriptome of these cultures under one specific growth condition. Hence, future transcriptome analysis should analyse these cultures under a variety of culture conditions (such as variations in light intensity or during nutrient depletion) to more fully encapsulate the metabolic potential of these organisms. Conversely, complete genome sequencing could be performed for each strain.

A combination of more genomic information and physiological studies to ascertain the function of hypothesised genes will enable the identification of more of the putative ORFs present in the current study. Furthermore, with increasing knowledge of the genomic capabilities of these organisms an ability to isolate into culture important representatives will be increased as culture media can be tailored to the specific requirements of individual taxa.

## **Chapter 7**

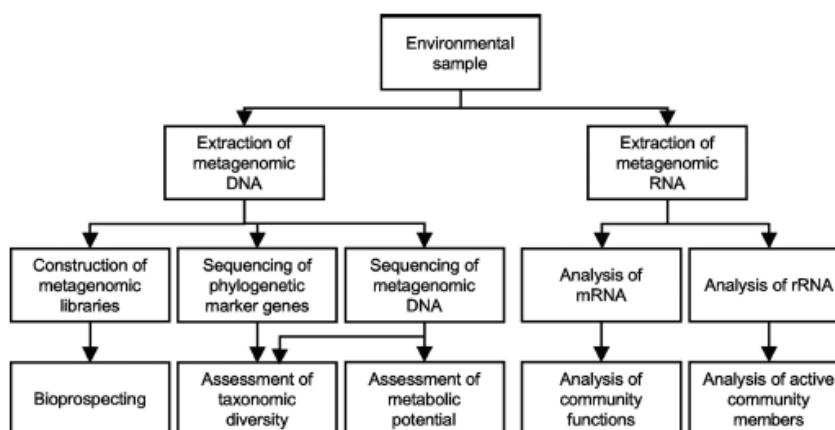
### **Metatranscriptomics of PPEs along an Atlantic Meridional Transect**

## **7.1 Introduction**

Microorganisms are integral to biogeochemical cycling in the oceans. An understanding of the roles of specific organisms in these cycles has previously relied on studies of pure cultures. However, with the limited ability to obtain into culture the majority of marine microbial species, culture independent genomic studies are needed to complement efforts to culture environmentally significant species (Riesenfeld *et al.*, 2004).

Two methods have been utilised to study the genomics of environmental samples depending on whether DNA or RNA is extracted (Figure 7.1). Metagenomics is the direct sequencing of environmental DNA and has proved a powerful tool for profiling the ecology and metabolism of complex communities (Simon and Daniel, 2011). The Sorcerer II Global Ocean Sampling Expedition (GOS) was a large scale metagenomic study of mainly marine surface waters (Rusch *et al.*, 2007). This extensive dataset consisting of 7.7 million sequencing reads (6.3 billion bp) significantly increased the knowledge of the diversity and phylogeny of protein families (Yooseph *et al.*, 2007). The GOS dataset, collected from more than 40 different sites with accompanying quantitative environmental features has enabled a comparative metagenomics approach to be undertaken. Comparisons between environments were able to show that processes such as photosynthesis, carbon and nitrogen fixation were strongly influenced by environmental variation (Gianoulis *et al.*, 2009). Recently, Cuvelier *et al.*, (2010) undertook a metagenomic study of pico-prymnesiophytes sorted by flow cytometry. Marine pico-prymnesiophytes are largely uncultured and this study was able to reveal some of the phylogenetic relationships and functional gene repertoires of this group. For example, this showed

that Fe-superoxide dismutases (SOD) were absent in the metagenome, but an alternative metal isoform containing nickel was present, suggesting a mechanism for the reduction in cellular iron demands of these organisms (Cuvelier *et al.*, 2010)



**Figure 7.1. Environmental genomic analysis of microbial communities from nucleic acids. (Reproduced from Simon and Daniel, 2011)**

Metagenomics provides information on the metabolic and functional potential of the microbial community (Figure 7.1). These DNA based studies hence give no information on those genes expressed, or not expressed, under a given set of environmental conditions meaning metagenomic studies cannot inform us of the metabolic ‘activity’ of the microbial community. To do that requires analysis of either gene expression (transcriptomics), protein profiles (proteomics) or individual assays of enzyme activity. Here, we chose a transcriptomics approach, specifically targeting the PPE fraction, to provide a molecular assessment of the potential regulatory factors controlling this group. This is based on the idea that transcriptional profiling will ‘let the organism inform us of the key environmental parameters that these organisms are responding to’. Such a molecular approach can potentially complement those ‘nutrient-addition bottle incubation’ studies that have often been

used to assess factors limiting growth rate/yield in specific microbial groups (e.g. see Moore *et al.*, 2008; Mills *et al.*, 2008; Davey *et al.*, 2008).

Until recently, the methods required for assessing gene expression in the environment have been challenging. mRNA only accounts for a small proportion of total RNA and generally has a short half life (Frias-Lopez *et al.*, 2008). Only very recently was the first microbial metatranscriptome study undertaken, focusing on the prokaryotic component at two aquatic sites (Poretsky *et al.*, 2005). Here, a relatively small metatranscriptomic library was constructed (sampling 400 clones) and included transcripts encoding proteins involved in sulphur oxidation, demonstrating proof of concept for the functional analysis of expressed genes from an environmental sample. Since then several other metatranscriptomic studies investigating marine prokaryotes have been undertaken (see Frias-Lopez *et al.*, 2008; Poretsky *et al.*, 2009). These have begun to include studies aimed at providing more quantitative analysis of the metatranscriptome (e.g. see Gifford *et al.*, 2011). In this latter work, at a coastal site off the southeastern USA, it was revealed that transcripts encoding ammonia and phosphate transporters were in high abundance, consistent with the elevated concentrations of these compounds at the survey site at the time of sampling, and highlighting the ability to correlate the expression of genes with the environmental conditions present. Metatranscriptomics have also been used to broaden our knowledge of prokaryotic non-coding RNAs (Shi *et al.*, 2009). Non-coding RNAs mostly function as regulators and are known from studies of a few model organisms to control environmentally significant processes such as amino acid biosynthesis and photosynthesis (see Steglich *et al.*, 2008; Vogel *et al.*, 2003). In contrast, far fewer eukaryotic metatranscriptomic libraries have been constructed,

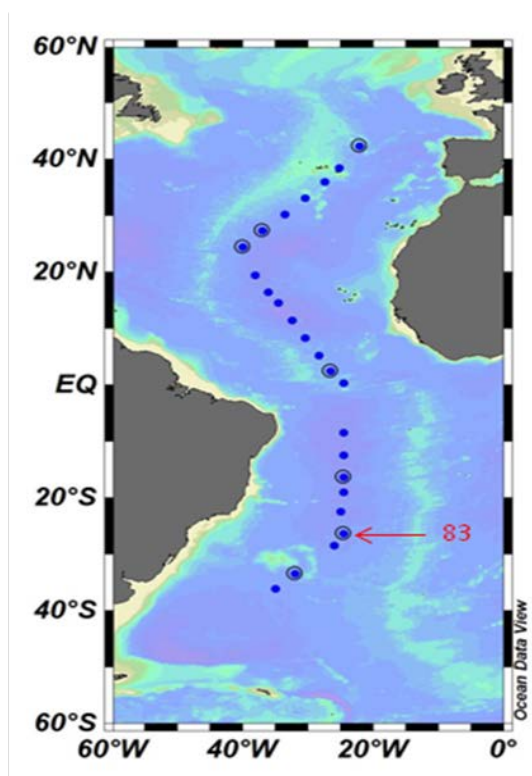
with appropriate methodologies generally lagging behind those developed for prokaryotes. This is despite the presence of a poly(A) tail on eukaryotic mRNAs that generally allows for simpler ‘enrichment’ of mRNA in the sample as well as eukaryotic mRNA being on average longer lived. The first example of a eukaryotic metatranscriptome was of geothermal algal mats and activated sludge (Grant *et al.*, 2006). More recently a metatranscriptomic library of eukaryotic plankton was created by John *et al.*, (2009). This study found that the eutrophic bay study site in Tampa Bay, Florida mainly comprised transcripts affiliated with diatoms, chlorophytes and dinoflagellates, and with transcripts encoding protein groups related to photosynthetic growth and various biogeochemical cycles readily identifiable. However, both of these studies targeted eukaryotes generally, i.e. encompassing heterotrophic, mixotrophic as well as photosynthetic lineages.

In this study the aim was to undertake a metatranscriptomic approach specifically targeting the PPE population (so as to exclude transcripts and hence ‘activities’ associated with heterotrophic lineages), using a flow sorting approach and developing a new pipeline for extracting and ‘amplifying’ mRNA from sorted cells. A PPE sample from a station along AMT18 in the south Atlantic gyre was used as a ‘test’ sample. It was envisaged that expression analysis of a natural PPE population would not only increase ‘functional’ knowledge of these organisms, which are critical to primary production (e.g. see Jardillier *et al.*, 2010), but would also allow future comparison of metatranscriptomes between samples so that we could begin to evaluate how the environment exerts differences in gene expression; in other words, and as already mentioned above, so that we can be informed ‘of the key environmental parameters that these organisms are responding to’.

## **7.2 Materials and Methods**

### **7.2.1 Sample collection**

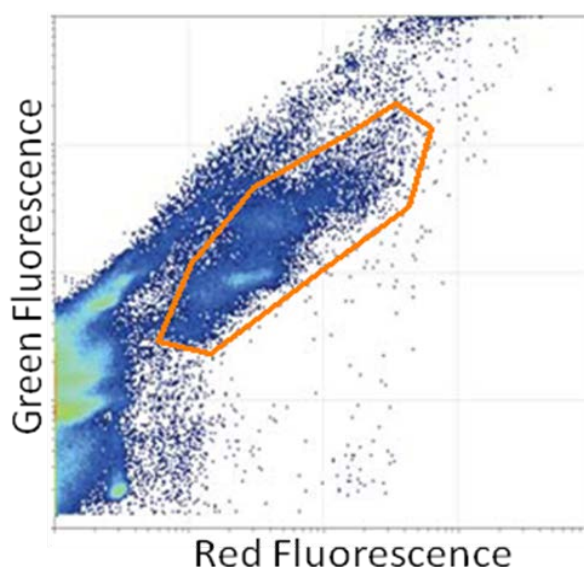
The seawater sample for metatranscriptome analysis was collected during the AMT18 cruise from station 83 (26.33°S, 24.59°W), 17m depth in the south Atlantic gyre (Figure 7.2) using a new sampling approach that allowed rapid concentration of cell material (see section 2.3.1.3). The concentrated seawater sample was immediately flash frozen and stored at -80°C before being transported back to the UK.



**Figure 7.2. The AMT 18 cruise highlighting the station sampled for metatranscriptomic analysis**

### 7.2.2 Flow cytometric sorting of the metatranscriptome seawater sample

The total PPE population (i.e. encompassing both Euk-A and Euk-B populations see chapter 4, Figure 7.3) of the southern gyre seawater sample was sorted using a MoFlo flow cytometer (Dako-Cytomation) using the highest sorting purity of the instrument and endogenous chlorophyll *a* fluorescence (FL3) and side scatter (SSC) as distinguishing factors (see section 2.3.1.4 for details). Cells were sorted directly into acid phenol and flash frozen before being stored at -80°C.



**Figure 7.3.** Flow cytometry scatterplot illustrating the picoeukaryote PPE fraction which was targeted for the metatranscriptome (adapted from Jardillier *et al.*, 2010).

### 7.2.3 Extraction of RNA and enrichment of mRNA

Total RNA was extracted from the flow cytometrically sorted cells using a Ribolysing technique specifically developed here (see section 2.3.2). mRNA was enriched from the total RNA sample using Terminator™ 5'-phosphate dependent exonuclease (see section 2.3.4). The enriched mRNA was amplified using Ambion's



MessageAmp™ II. Further amplification of the resulting cDNA was achieved by the addition of a poly(A) tail and amplification using an Oligo (dT) primer as described in section 2.3.6.

#### **7.2.4 Sequencing of the metatranscriptomic sample**

cDNA (~ 5 ug) was sent to the NERC Biomolecular Analysis Facility in Liverpool. The cDNA was sequenced using a Next Generation Roche 454 FLX sequencer.

#### **7.2.5 Analysis of 454 reads**

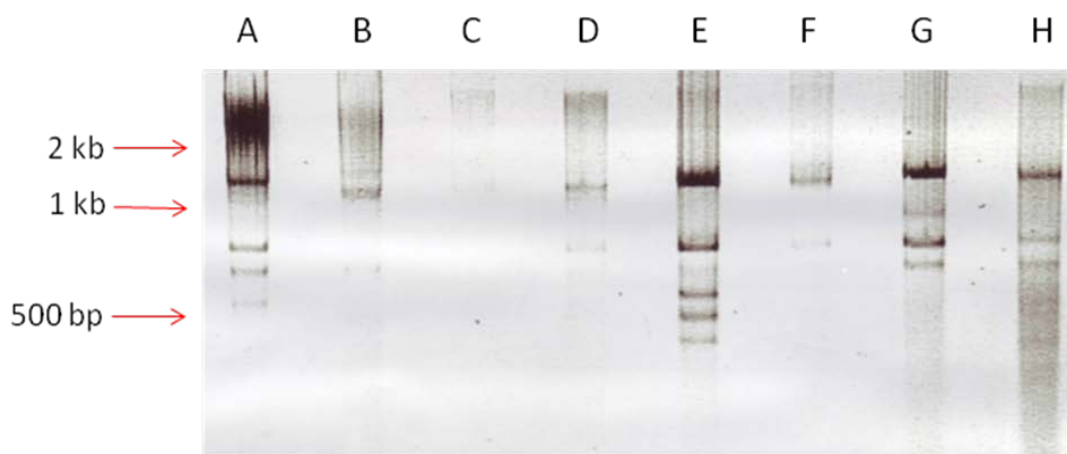
Contigs from the raw reads were assembled using the Newbler program (Roche). These contigs were analysed against NCBI protein nr database using BLASTx. Phylogenetic analysis of the top BLASTx reads was undertaken using the MEGAN program (see 2.4.3.6). Contigs were submitted to RAMMCAP on the CAMERA website (<http://camera.calit2.net/>) as described in section 2.4.3.3. The contigs were also analysed against the Clusters of Orthologous Genes for eukaryotes (KOG) and Kyoto Encyclopedia of genes and genomes (KEGG) databases for functional annotation (see 2.3.4.5 and 2.3.4.5). Prediction of the localisation of transcripts was undertaken using TargetP (Emanuelsson *et al.*, 2007). This program identifies transit peptides of nuclear encoded genes to the chloroplast and mitochondrion (<http://www.cbs.dtu.dk/services/TargetP/>).

### **7.3 Results**

#### **7.3.1 Pipeline for the extraction and amplification of environmental RNA**

RNA was extracted from various cultures as a proof that RNA could be extracted from a variety of phytoplankton. A chrysophyte, prasinophyte and bacillariophyte

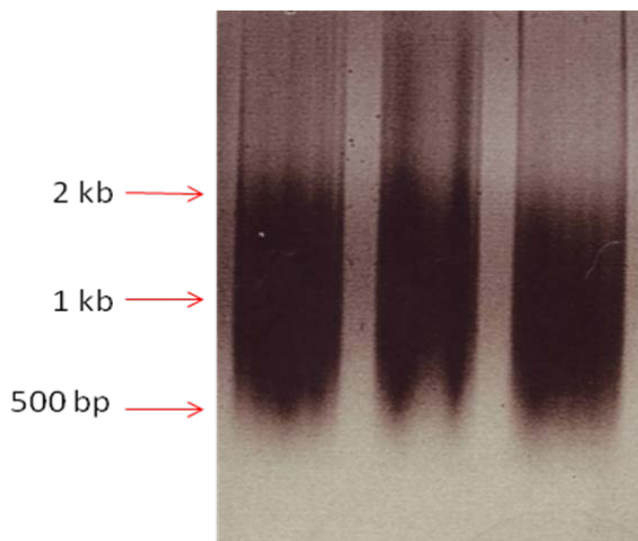
were extracted along with a variety of cultures obtained during the AMT 18 cruise (Figure 7.4).



**Figure 7.4. Agarose gel electrophoresis of RNA extracted from a variety of phytoplankton cultures; A) AMT station 97 - surface, B) Chrysophyte sp., C) AMT 56 - surface, D) Bacillariophyte sp., E) AMT station 32 – 124m, F) AMT station 16 – 129m, G) AMT station 19 – 132m, H) Prasinophyte sp. Arrows indicate the size of markers.**

The extracted RNA from flow sorted PPE cells obtained from station 83 17m depth, a southern gyre station. The RNA was amplified using MessageAmpII™ resulting in ~ 1500 ng of aRNA which was converted into cDNA using random primers. The cDNA was run on a 1% (w/v) agarose gel and cDNA between 500 bp and 2 kb was extracted from the gel and used as a template for further amplification.

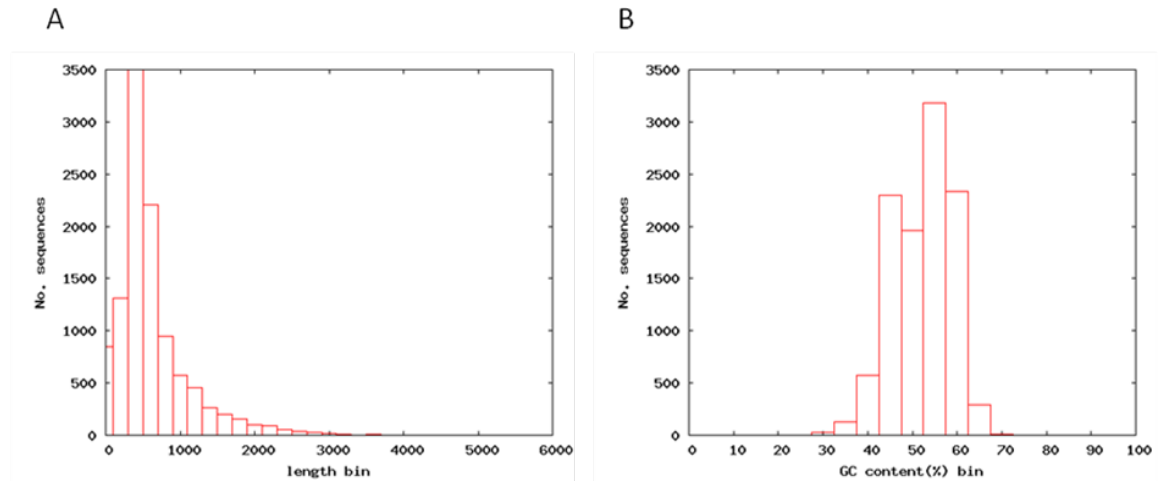
The final amplification step using poly (dT) tailed primers led to an increase in the amount of cDNA produced. The cDNA was run on a 1.5% agarose gel and cDNA between 500 bp and 2 kb was extracted from the gel (Figure 7.5). The size fractionated cDNA (5 µg) was sent for sequencing.



**Figure 7.5. PCR products following the amplification of cDNA derived from the southern gyre station 83 17m depth along AMT18, showing the size range of products sent for sequencing.**

### **7.3.2 General analysis of the southern gyre metatranscriptome sequences**

Sequencing of the southern gyre AMT18 metatranscriptome sample produced a total of 442,913 reads. The average length of the reads was 380 bp with a range between 20 - 881 bp. The Newbler work flow was used to assemble contigs. A total of 10,802 contigs were formed with an average length of 714 bp. The length of the contigs ranged from 100 - 5223 bp (Figure 7.6A). The median binned %GC content of the contigs was 55% (Figure 7.6B).



**Figure 7.6. A) The average length of contigs produced from the southern gyre AMT18 transcriptome sample and B) the binned %GC content of the transcriptome.**

**Table 7.1: Phylogenetic affiliation of eukaryotic rRNA transcripts in the metatranscriptome**

Contig	Top BLAST hit	E value	Top culture hit	% identity to culture
contig00246	Uncultured marine eukaryote	3E-72	<i>Nitzschia thermalis</i>	100
contig00363	Uncultured marine eukaryote	1E-159	<i>Nitzschia thermalis</i>	99
contig00642	<i>Galeidinium rugatum</i>	2E-65	<i>Galeidinium rugatum</i>	97
contig00748	uncultured diatom	2E-41	<i>Cymbella cistuliformis</i>	97
contig01023	Uncultured marine eukaryote	2E-87	<i>Prorocentrum glenanicum</i>	96
contig01608	<i>Akashiwo sanguinea</i>	0E+00	<i>Akashiwo sanguinea</i>	97
contig01975	<i>Peridinium foliaceum</i>	0E+00	<i>Peridinium foliaceum</i>	97
contig03965	<i>Gymnodiniaceae</i> sp. AW-2009	6E-56	<i>Gymnodiniaceae</i> sp.	88
contig04034	<i>Asterionella formosa</i>	3E-62	<i>Asterionella formosa</i>	100
contig04110	<i>Kryptoperidinium foliaceum</i>	3E-164	<i>Kryptoperidinium foliaceum</i>	100
contig04371	<i>Phaeodactylum tricornutum</i>	0E+00	<i>Phaeodactylum tricornutum</i>	97
contig04800	<i>Akashiwo sanguinea</i>	1E-62	<i>Akashiwo sanguinea</i>	96
contig05404	<i>Gymnodinium impudicum</i>	4E-32	<i>Gymnodinium impudicum</i>	96
contig05502	<i>Gymnodinium aureolum</i>	4E-71	<i>Gymnodinium aureolum</i>	90
contig05625	<i>Cylindrotheca closterium</i>	6E-50	<i>Cylindrotheca closterium</i>	100
contig05813	Uncultured eukaryote	0E+00	<i>Galeidinium rugatum</i>	94
contig06069	<i>Cylindrotheca closterium</i>	2E-95	<i>Cylindrotheca closterium</i>	93
contig06325	<i>Phaeodactylum tricornutum</i>	2E-90	<i>Phaeodactylum tricornutum</i>	100

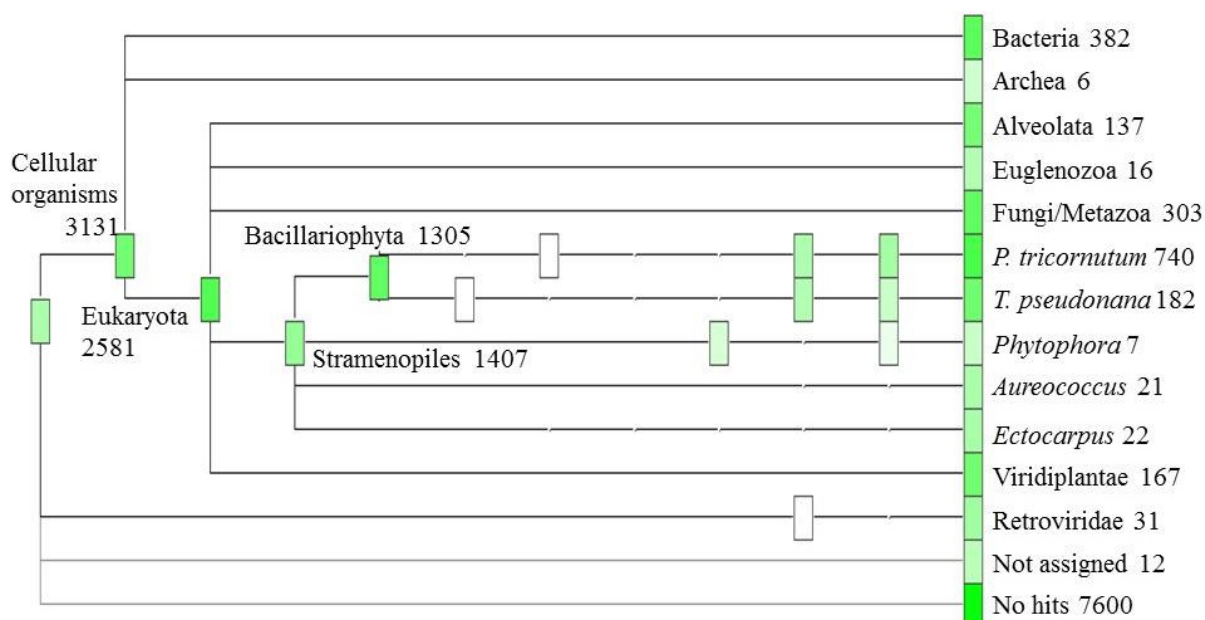
**Table 7.1 cont.**

Contig	Top BLAST hit	E value	Top culture hit	% identity to culture
contig06401	<i>Nitzschia</i> sp	9E-157	<i>Nitzschia</i> sp	100
contig06431	Uncultured alveolate	4E-140	<i>Takayama cf. pulchellum</i>	98
contig06891	<i>Kryptoperidinium foliaceum</i>	0E+00	<i>Kryptoperidinium foliaceum</i>	98
contig06898	<i>Durinskia baltica</i>	0E+00	<i>Durinskia baltica</i>	98
contig07030	<i>Phaeodactylum tricornutum</i>	2E-79	<i>Phaeodactylum tricornutum</i>	100
contig07854	Uncultured marine eukaryote	5E-108	<i>Cymbella inflata</i>	100
contig08233	<i>Galeidinium rugatum</i>	0E+00	<i>Galeidinium rugatum</i>	94
contig08350	<i>Cymbella bipartita</i>	0E+00	<i>Cymbella bipartita</i>	99
contig08531	<i>Kryptoperidinium foliaceum</i>	1E-138	<i>Kryptoperidinium foliaceum</i>	93
contig08534	<i>Kryptoperidinium foliaceum</i>	2E-142	<i>Kryptoperidinium foliaceum</i>	94
contig08730	<i>Alexandrium affine</i>	5E-134	<i>Alexandrium affine</i>	94
contig08923	Uncultured marine eukaryote	0E+00	<i>Nitzschia cf. pusilla</i>	100
contig09312	<i>Galeidinium rugatum</i>	2E-133	<i>Galeidinium rugatum</i>	98
contig09393	<i>Nitzschia</i> sp. MD1	3E-137	<i>Nitzschia</i> sp. MD1	98
contig09716	Uncultured marine alveolate	1E-118	<i>Karlodinium veneficum</i>	100
contig09800	<i>Kryptoperidinium foliaceum</i>	9E-103	<i>Kryptoperidinium foliaceum</i>	100
contig09802	Uncultured marine eukaryote	1E-100	<i>Prorocentrum glenanicum</i>	97
contig10115	<i>Bacillaria paxillifer</i>	1E-78	<i>Bacillaria paxillifer</i>	100
contig10126	<i>Cylindrotheca closterium</i>	1E-79	<i>Cylindrotheca closterium</i>	97
contig10216	<i>Kryptoperidinium foliaceum</i>	3E-38	<i>Kryptoperidinium foliaceum</i>	92
contig10262	<i>Nitzschia</i> sp. MD1	5E-68	<i>Nitzschia</i> sp. MD1	96
contig10416	<i>Goussia</i> sp. BMR-2011	4E-58	<i>Goussia</i> sp. BMR-2011	96
contig10421	<i>Kryptoperidinium foliaceum</i>	2E-60	<i>Kryptoperidinium foliaceum</i>	97
contig10477	<i>Navicula perminuta</i>	1E-62	<i>Navicula perminuta</i>	100
contig10534	Uncultured eukaryote clone	2E-58	<i>Pentapharsodinium</i> sp	98
contig10759	<i>Kryptoperidinium foliaceum</i>	3E-43	<i>Kryptoperidinium foliaceum</i>	100
contig10774	<i>Prorocentrum consutum</i>	1E-29	<i>Prorocentrum consutum</i>	92

### 7.3.3 Phylogenetic analysis of the metatranscriptome

Out of the 10,802 contigs in the sample only 67, representing 0.62%, were related to rRNA reads. Phylogenetic analysis of the rRNA reads showed that the metatranscriptome was dominated by two eukaryotic classes: Dinophyceae and Bacillariophyceae (Table 7.1).

The phylogeny of the top BLASTx hits showed that the vast majority of the contigs (70%) had no hits against the NCBI nr database. Eukaryota accounted for 24% of the contigs with bacteria accounting for 3.5%. The dominant group in the Eukaryota was the stramenopiles accounting for 54% (13% of all contigs) of the eukaryotic reads. Smaller contributions were made by the Alveolata (5.3% (1.3% of all contigs)), Viridiplantae (6.4% (1.5% of all contigs)) and fungi/metazoan (11.7% (2.8% of all contigs)). Within the stramenopiles most hits were against the sequenced diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* (Figure 7.7).



**Figure 7.7. Phylogenetic analysis of the BLASTx top hits from the AMT18 southern gyre metatranscriptome.**

### 7.3.4 Analysis of the AMT18 southern gyre metatranscriptome against the KOG database

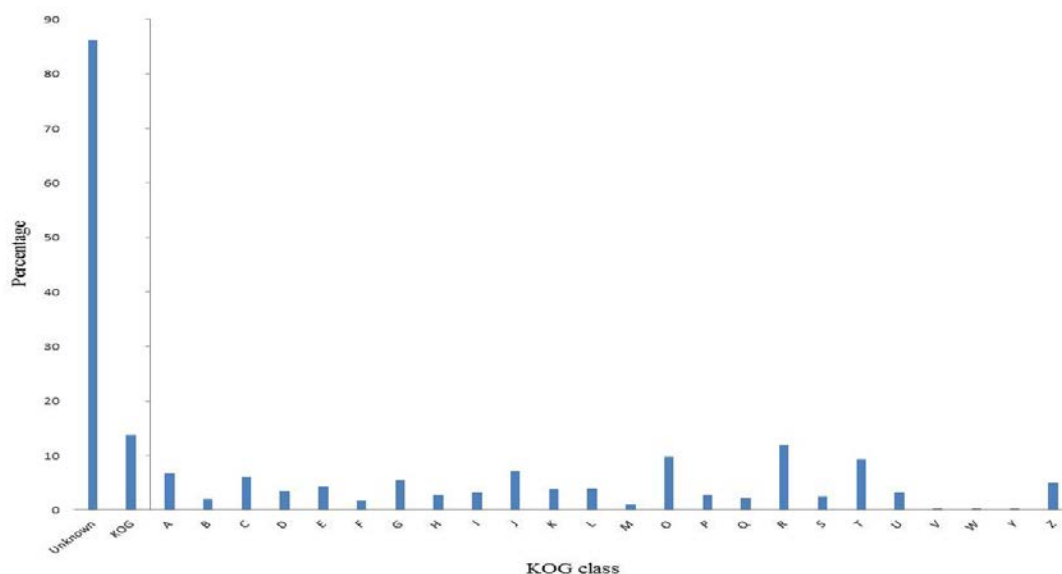
The number of predicted ORFs in the metatranscriptome was 8,118. The average length of the ORFs was 106 aa with a range between 20 – 895 aa. Out of the 8,118

predicted ORFs there were 1118 hits against the KOG database (13.8%). A significant fraction (12% (1.6% of total ORFs)) of the ORFs which were predicted against the KOG database could only be assigned to a general function (see Figure 7.8). KOG class O (Post-translational modification, protein turnover, chaperones) accounted for 9.7% (1.3% of total ORFs) of the assigned ORFs.

Classes T (Signal transduction mechanisms) and J (Translation, ribosomal structure and biogenesis) accounted for 9.3% (1.3% of total ORFs) and 7.1% (1.0% of total ORFs), respectively of the hits against the KOG database. Classes A (RNA processing and modification) and C (Energy production and conversion) accounted for 6.9% (0.9% of total ORFs) and 6.0% (0.8% of total ORFs) of the hits. The remaining KOG classes only accounted for minor fractions of the hits (Figure 7.8). The number of ORFs predicted against the PFAM database was 991 with only 412 ORFs hitting the TIGRFAM database.

### **7.3.5 Light harvesting and carbon fixation**

Transcripts encoding enzymes responsible for  $C_4$  photosynthesis were found in the metatranscriptome. Thus, the metatranscriptome contained contigs encoding phosphoenolpyruvate carboxylase (PEPCase), phosphoenolpyruvate carboxykinase (PEPCKase), and pyruvate orthophosphate dikinase (PPDK) (Table 7.2). The transcriptome also shows a putative read for a phosphoenolpyruvate translocator although no transit peptides could be identified in the transcript in order for cellular localisation to be predicted using TargetP.



**Figure 7.8.** The percentage of ORFs classified as KOGs and percentages assigned to each of the KOG classes. Classes are: A- RNA processing and modification; B - Chromatin structure and dynamics; C - Energy production and conversion; D - Cell cycle control, cell division, chromosome partitioning; E - Amino acid transport and metabolism; F - Nucleotide transport and metabolism; G - Carbohydrate transport and metabolism; H - Coenzyme transport and metabolism; I - Lipid transport and metabolism; J - Translation, ribosomal structure and biogenesis; K - Transcription; L - Replication, recombination and repair; M - Cell wall/membrane/envelope biogenesis; O - Posttranslational modification, protein turnover, chaperones; P - Inorganic ion transport and metabolism; Q - Secondary metabolites biosynthesis, transport and catabolism; R - General function prediction only; S - Function unknown; T - Signal transduction mechanisms; U - Intracellular trafficking, secretion, and vesicular transport; V - Defence mechanisms; W - Extracellular structures; Y - Nuclear structure; Z - Cytoskeleton.

A single contig related to RuBisCO (contig02229). The phylogenetic relationship of this RuBisCO was to a member of the Dinophyceae (*Galeidinium rugatum*).

BLASTx analysis of the contigs revealed that transcripts encoding fucoxanthin chlorophyll *a/c* binding proteins were the only light harvesting complex binding

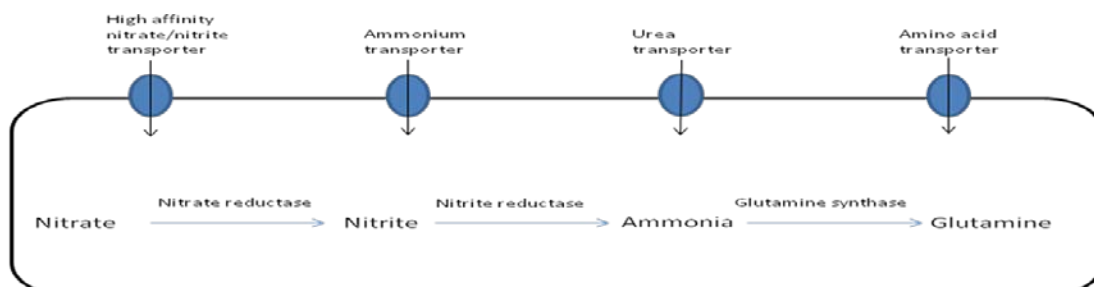


proteins observed. Thus there were no contigs relating to chlorophyll *a/b* binding proteins present.

Transcripts encoding two types of photoreceptor were also observed (Table 7.3). The top BLAST hit for the phytochrome transcript was related to the diatom *P. tricornatum* whilst the highest hit identified as a cryptochrome was related to *Spirosoma* with two hits having greater e-values relating to *P. tricornatum* predicted proteins. Cryptochromes, which function as blue light receptors and phytochromes which perceive red/far red light were found. No contigs relating to rhodopsin or phototropin, which detect green light, were detected.

### 7.3.6 Nutrient assimilation and acquisition

Contigs encoding enzymes required for the assimilation of nitrogen were also detected. These included transcripts encoding nitrate reductase, nitrite reductase and glutamine synthase. Also present in the metatranscriptome were a variety of transcripts encoding nitrogen transporters including those encoding a Major Facilitator Superfamily (MFS) transporter for nitrate/nitrite and transporters for ammonia, urea and amino acids (Figure 7.9 and Table 7.4).



**Figure 7.9. Transporters and enzymes associated with nitrogen acquisition and assimilation present in the southern gyre metatranscriptome.**

**Table7.2. Transcripts involved in carbon fixation and the phylogenetic affiliation**

Contig	Hit	Phylogeny of top BLAST hit	E value	% positives	Length of hit
2150	malate dehydrogenase (oxaloacetate-decarboxylating)(NADP+)	<i>Gramella forsetii</i>	1E-98	85	287
1082	glyceraldehyde-3-phosphate dehydrogenase (NAD(P))	<i>Phaeodactylum tricornutum</i>	1E-169	96	200
1318		<i>Phaeodactylum tricornutum</i>	5E-138	88	330
2857		<i>Scrippsiella trochoidea</i>	2E-119	97	229
4836		<i>Scrippsiella trochoidea</i>	4E-104	96	201
5439		<i>Alexandrium fundyense</i>	1E-75	92	163
6059		<i>Lepidodinium chlorophorum</i>	9E-90	96	178
9134		<i>Scrippsiella trochoidea</i>	3E-49	95	106
9283		<i>Scrippsiella trochoidea</i>	1E-44	95	99
9556		<i>Gonyaulax polyedra</i>	6E-23	98	56
485	aspartate aminotransferase	<i>Solibacter usitatus</i>	2E-71	66	355
7424		<i>Phaeodactylum tricornutum</i>	3E-47	76	143
5355	phosphoribulokinase	<i>Phaeodactylum tricornutum</i>	8E-85	94	191
831	pyruvate kinase	<i>Phaeodactylum tricornutum</i>	2E-86	90	191
4919		<i>Rhodothermus marinus</i>	7E-56	74	200
531	phosphoglycerate kinase	<i>Kryptoperidinium foliaceum</i>	2E-77	93	185
3844		<i>Phaeodactylum tricornutum</i>	3E-108	95	228
6073		<i>Phaeodactylum tricornutum</i>	1E-48	95	108
5767	pyruvate,orthophosphate dikinase	<i>Thalassiosira pseudonana</i>	6E-33	77	120
2259	phosphoenolpyruvate carboxylase	<i>Thalassiosira pseudonana</i>	3E-121	85	300
7308	phosphoenolpyruvate carboxylase	<i>Thalassiosira pseudonana</i>	2E-27	62	156
2229	ribulose-bisphosphate carboxylase large chain	<i>Galeidinium rugatum</i>	1E-168	97	305
2580	phosphoenolpyruvate carboxykinase (ATP)	<i>Pirellula staleyi</i>	7E-47	70	171
97	fructose-bisphosphate aldolase, class II	<i>Phaeodactylum tricornutum</i>	4E-161	91	270
1116		<i>Heterocapsa triquetra</i>	7E-83	91	175
3440		<i>Heterocapsa triquetra</i>	3E-93	92	120
7533		<i>Pfiesteria piscicida</i>	5E-87	97	161
8869		<i>Pfiesteria piscicida</i>	9E-34	97	70
2796	ribulose-phosphate 3-epimerase	<i>Phytophthora infestans</i>	4E-59	68	221
3075	triosephosphate isomerase	<i>Phaeodactylum tricornutum</i>	2E-82	82	207

**Table 7.3. Blast results for transcripts relating to photoreceptors and the phylogenetic affiliation.**

Contig	Hit	Species	E value	% positives	Length of hit
2682	Cryptochrome	<i>Spirosoma</i>	1.00E-11	46	184
4164	Phytochrome	<i>P.tricornatum</i>	1.00E-69	78	212
455	PAS sensor protein	<i>Desulfobacterium</i>	1.00E-58	58	330
2308	PAS sensor protein	<i>Glaciecola</i> sp.	4.00E-52	65	265
3449	PAS sensor protein	<i>Syntrophomonas</i>	3.00E-37	68	186

The southern gyre metatranscriptome also contained transcripts encoding proteins involved in the acquisition and assimilation of sulphate. These included sulphate reductase, sulphate oxidase, sulphate adenylyltransferase and the sulphate permease (SulP). However transcripts encoding cysteine synthase were not detected.

### 7.3.7 Utilisation of iron in the metatranscriptome

Transcripts encoding cytochrome  $c_6$ , encoding an electron carrier between the cytochrome  $b_6/f$  complex and photosystem I, were also detected in the sample. The alternative carrier, plastocyanin, was not detected in the metatranscriptome. The metatranscriptome also contained transcripts encoding both ferredoxin and flavodoxin, two high affinity enzymes used in the assimilation of iron: ferric reductase and iron permease and contigs encoding an ATP dependent ABC transporter specific to iron (Table 7.5).

**Table 7.4 Transcripts identified in the metabolism and transport of nitrogen.**

Contig	Transcript	Phylogeny of top hit	E value	% positives	Length of hit
Nitrogen metabolism					
contig00825	Glutamate synthase (ferredoxin)	<i>Psychroflexus torquis</i>	4E-151	72	478
contig01219	Nitrate reductase (NADH)	<i>Cylindrotheca fusiformis</i>	1E-56	71	98
contig00500	Ferredoxin-nitrite reductase	<i>Thalassiosira pseudonana</i>	3E-126	77	357
contig05318	Aminomethyltransferase	<i>Ectocarpus siliculosus</i>	1E-21	56	132
contig00051	Formamidase	<i>Roseiflexus</i>	2E-69	60	283
contig01193	Nitrilase	<i>Thalassiosira pseudonana</i>	2E-98	74	303
contig00394	Aspartate ammonia-lyase	<i>Hahella chejuensis</i>	6E-51	79	95
contig00104	Histidine ammonia-lyase	<i>Marivirga tractuosa</i>	1E-142	95	287
contig05517	Glutamine synthetase	<i>Escherichia coli</i>	5E-42	100	82
contig06206		<i>Escherichia coli</i>	8E-96	98	177
contig00257		<i>Phaeodactylum tricornutum</i>	0E+00	88	449
Transporters					
contig03491	Urea transporter	<i>Phaeodactylum tricornutum</i>	1E-06	73	56
contig00180	Amino acid transporter	<i>Phytophthora infestans</i>	4E-34	55	260
contig00454	Cationic amino acid transporter	<i>Ochrobactrum intermedium</i>	2E-25	46	334
contig01215	Amino acid transporter	<i>Plesiocystis pacifica</i>	5E-19	47	357
contig02028	Amino acid/polyamine transporter	<i>Thalassiosira pseudonana</i>	6E-05	52	98
contig02554	neutral amino acid transporter	<i>Phytophthora infestans</i>	3E-13	60	110
contig04242	proton/glutamate symporter	<i>Salinibacter rubber</i>	2E-17	73	133
contig04909	Amino acid/polyamine transporter	<i>Thalassiosira pseudonana</i>	6E-20	52	170
contig09030	Amino acid permease	<i>Escherichia coli</i>	5E-56	98	108
contig00237	nitrate transporter	<i>Symbiodinium</i>	4E-91	79	284
contig00834	MFS nitrate nitrite transporter	<i>Symbiodinium</i>	2E-168	85	490
contig02151	AMT ammonium transporter	<i>Perkinsus marinus</i>	4E-30	62	167
contig04818	AMT ammonium transporter	<i>Perkinsus marinus</i>	2E-20	62	144

**Table 7.5. Utilisation and transport of iron in the AMT 18 metatranscriptome**

Contig	Hit	Phylogeny of top hit	E value	% positives	Length of hit
contig00587	Cytochrome c <sub>6</sub>	<i>Chaetoceros gracilis</i>	4E-43	82	131
contig03200	Ferredoxin	<i>Rhodospirillum centenum</i>	9E-21	70	99
contig06112	Ferredoxin	<i>Thalassiosira pseudonana</i>	6E-26	88	48
contig00946	Flavodoxin	<i>Aquifex aeolicus</i>	1E-36	76	131
contig07289	Flavodoxin	<i>Idiomarina baltica</i>	5E-19	81	68
contig03130	Iron permease	<i>Aureococcus anophagefferens</i>	5E-05	40	262
contig01869	Ferric reductase	<i>Tetrahymena thermophila</i>	3E-25	53	287
contig02087	Ferric reductase	<i>Phytophthora infestans</i>	4E-04	54	94
contig02440	Ferric iron transport (ABC)	<i>Psychroflexus torquis</i>	3E-41	79	124

## **7.4 Discussion**

### **7.4.1 Enrichment of mRNA in the metatranscriptome**

RNA from sorted PPE cells extracted using a Ribolyser was subjected to mRNA enrichment using Terminator<sup>TM</sup> 5'-phosphate dependent exonuclease (Epicentre) (see section 2.3.4). The percentage of reads relating to rRNA in the metatranscriptome was 0.62%. This compares very favourably to the use of magnetic beads for the enrichment step (see section 6.4.1; Grant *et al.*, 2006; John *et al.*, 2009), an approach not appropriate with sorted cells due to the small amount of starting material available. Indeed, this exonuclease approach resulted in a smaller percentage of rRNA reads compared with either the John *et al.*, 2009 or Grant *et al.*, 2006 studies (0.62% versus 2% or 4%, respectively). Thus, eukaryotic mRNA enrichment using the exonuclease approach clearly dramatically reduces the amount of rRNA present in the sample and appears to be a method of choice when only a small amount of starting material is available.

### 7.4.2 Phylogenetics of the metatranscriptome

Targeted metatranscriptomic analysis of eukaryotic picophytoplankton is clearly at an early stage. Few studies have been undertaken to assess the functional capabilities of these important organisms in the marine environment.

The rRNA reads found in this study were mainly related to Dinophyceae and Bacillariophyceae with a single contig related to Trebouxiophyceae. The domination of these classes in the metatranscriptome was in contrast to the plastid biased 16S rRNA clone library reported in chapter 4 (section 4.3.3), constructed from the same seawater sample. In the plastid biased 16S rRNA clone library Prymnesiophyceae and Trebouxiophyceae were the dominant groups with a small contribution from members of the Chrysophyceae. The discrepancies between these results could be down to several factors. Firstly, the plastid biased 16S rRNA clone library was constructed specifically on the Euk-A PPE population while the metatranscriptomic library comprised both Euk-A and Euk-B populations. Secondly, all the rRNA reads obtained in the metatranscriptomic library were of nuclear origin. As shown previously (section 4.3.8) different results were obtained from the nuclear and plastid rRNA clone libraries presumably due to the different PCR biases associated with these primer sets (e.g. see Shi *et al.*, 2011). Here, the bias may be more a result of the differing copy numbers of rRNA genes present in these PPE classes (see Zhu *et al.*, 2005) or a combination of the above factors could explain the differences observed between the rRNA reads in the metatranscriptome compared with that of the clone libraries.

The phylogeny of the BLASTx reads showed that of the reads that could be assigned to a taxon, the majority were related to stramenopiles and in particular

Bacillariophyceae. Only a small numbers of reads were assigned to Alveolata and Viridiplantae. The dominance of Bacillariophyceae in the BLASTx top hits, with a relatively minor contribution associated with Dinophyceae, may be due to the bias in the GenBank database towards organisms which have been heavily studied or have had their genomes sequenced. Two diatom genomes, *T.pseudonana* (Armbrust *et al.*, 2004) and *P. tricornutum* (Bowler *et al.*, 2008), have been completed but no Dinophyceae genome is currently available. The absence of the latter is presumably a result of the enormous genome sizes predicted for dinoflagellates (see Hackett *et al.*, 2006). It is possible that a large proportion of the sequences with no hits show homology to genes from organisms which have yet to be characterised, or, of genomes yet to be submitted to Genbank. This could explain the paucity of hits against classes such as Chrysophyceae which have been poorly studied thus far, especially those from marine systems. As more genomes become available, our increase in knowledge will lead to more sequences from a metatranscriptomic library being able to be assigned and subsequently annotated.

The median binned %GC content of the metatranscriptome was 55%. This is similar to the GC content (53.7%) of the coding sequences of the diatom *P. tricornutum* (Scala *et al.*, 2002). However, given that the %GC content of the metatranscriptome ranges between 30-70% it is likely that transcripts arising from other lineages are contained within it. Moreover, with the absence of sequenced genomes for relevant taxa it is unclear how this average %GC content will map onto this.

### 7.4.3 Functional prediction of contigs

In the metatranscriptome there were 8,118 ORFs predicted by metagene. However, only 11.8% of these predicted ORFs could be assigned a function against the KOG database (see section 7.3.3). The KOG database contains protein clusters from the complete genomes of seven eukaryotes including the higher plant *Arabidopsis thaliana* (Tatusov *et al.*, 2003). The addition of new eukaryotic genomes to the KOG database is expected to result in a substantial increase in the coverage of eukaryotic genomes with KOGs. Within the reads which were assigned a KOG, the largest class was R (General function). Currently, functional analysis of metatranscriptomic libraries is limited by the ability to assign the predicted ORFs to the relevant KOG class. Certainly, analysis of marine metatranscriptomic samples would be aided by the addition of sequenced genomes, which are environmentally relevant, to the KOG database. As more genomes become available and annotation of genes is undertaken the ability to identify genes in metatranscriptomic samples is likely to be increased considerably.

### 7.4.4 Light harvesting and carbon fixation: a metatranscriptomics perspective

Carbon dioxide is vital for photosynthesis, and marine phytoplankton have developed mechanisms to alleviate carbon limitation including metabolic adaptations and mechanisms to concentrate CO<sub>2</sub> at the active site of RuBisCO (Moroney and Somanchi, 1999). Similar to what was seen for the prasinophyte clade VIIA and *Ochromonas* cDNA libraries (see section 6.4.2) the metatranscriptomic library also included transcripts encoding enzymes involved in C<sub>4</sub> photosynthesis (section 7.3.5), including phosphoenolpyruvate carboxykinase, phosphoenolpyruvate carboxylase and pyruvate-phosphate dikinase. The presence of a C<sub>4</sub> photosynthesis/CCM



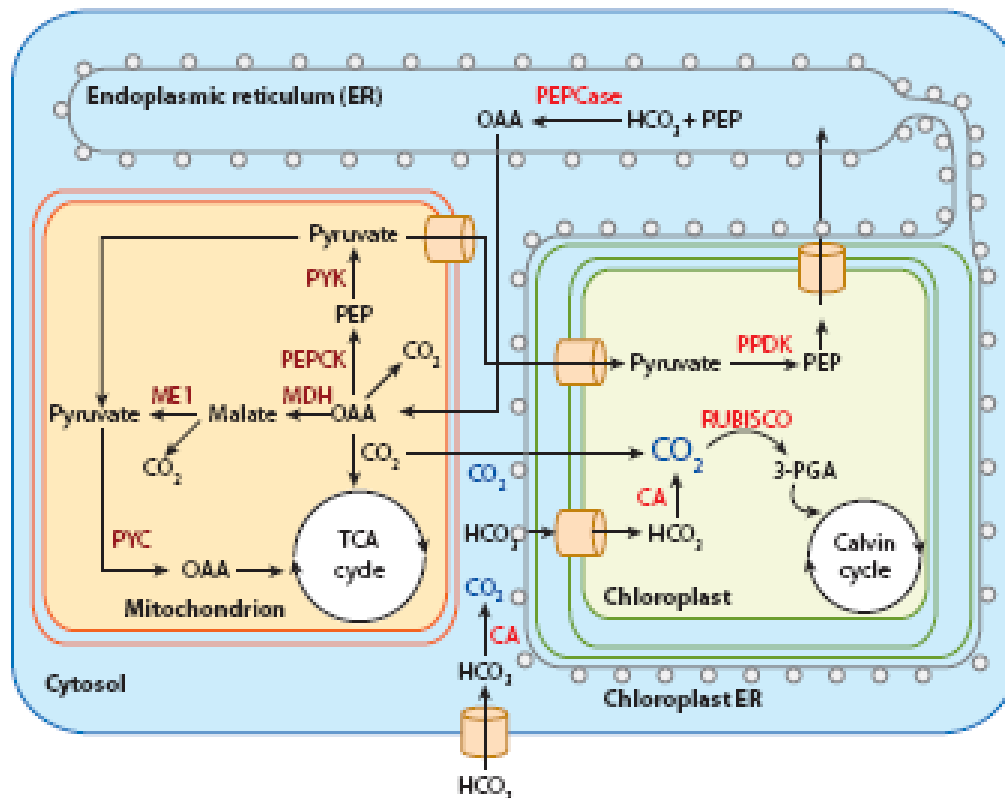
mechanism has been proposed in other unicellular marine algae with the appropriate suite of genes present in the genomes of *Ostreococcus*, *Micromonas* and the diatoms *T. pseudonana* and *P. tricornutum* (Parker *et al.*, 2008). The presence of transcripts encoding phosphoenolpyruvate carboxykinase suggests that a pathway more akin to that proposed in diatoms (Reinfelder *et al.*, 2004) than the pathways proposed in *Ostreococcus/Micromonas* is present in the southern gyre metagenomes (Worden *et al.*, 2009). The proposed mechanism of C<sub>4</sub> photosynthesis in diatoms includes the conversion of HCO<sub>3</sub><sup>-</sup> and phosphoenolpyruvate (PEP) into oxaloacetate (OAA) in the endoplasmic reticulum. OAA is hypothesised to be transported into the mitochondrion where it is converted into malate by malate dehydrogenase (see Figure 7.10). Malate is then converted into pyruvate, releasing CO<sub>2</sub> into the mitochondria. The diatom genome has no OAA and malate transporters in the plastid and it is hypothesised that CO<sub>2</sub> is concentrated in the mitochondrion and then delivered into the plastid for fixation by RuBisCO (Figure 7.10). No malate or OAA transporters were detected in the metatranscriptome suggesting that a pathway similar to that hypothesised in diatoms may be the dominant pathway in the transcriptome.

Transcripts affiliated to phytochrome, a red light photoreceptor (Sharrock, 2008) and cryptochrome, which function as blue-light photoreceptors (Cashmore *et al.*, 1999), were also found in the southern gyre metatranscriptome (see section 7.3.4). However, no significant matches of transcripts to rhodopsin or phototropins were present in the metatranscriptome. The presence of transcripts encoding these photoreceptors in the metatranscriptome suggests that organisms present in the sample could perceive red and blue light but not green light, perhaps due to the fact

that green light persists to depth only in coastal waters (Montsant *et al.*, 2007). Consistent with the phylogenetic affiliation of BLASTx reads mainly to members of the stramenopiles and especially diatoms, it is noteworthy that sequenced diatom genomes also contain genes encoding phytochrome and cryptochrome (Armbrust *et al.*, 2004; Bowler *et al.*, 2008). However, the role of phytochromes in marine algae is still unclear, especially due to the high attenuation of red light in water. In higher plants the detection of the Red:Far Red ratio represents the start and end of the photoperiod or shading under canopies (Ragni and D'Alcala, 2004). Phototaxis has been reported in the cyanobacteria to be negative in response to red light (Kondou *et al.*, 2001). This behaviour could be due to the avoidance of supersaturating irradiance regimes at the surface. However, in green algae red light has been reported to have a role in swimming velocity and chloroplast orientation with the suggestion that these organisms are trying to exploit favourable light conditions (Ragni and D'Alcala, 2004). A reduction in the blue: red light ratio is known to occur at high chlorophyll *a* concentrations. The information obtained from this ratio is hypothesised to be involved in the triggering and timing of processes such as aggregation and resting strategies (Ragni and D'Alcala, 2004).

Transcripts encoding only one type of light harvesting complex binding protein were present in the metatranscriptome (section 7.3.4), i.e. those encoding the fucoxanthin chlorophyll *a/c* binding protein. Such binding proteins are typical of diatoms and chrysophytes (Andersen *et al.*, 1999). However, the absence of transcripts encoding chlorophyll *a/b* binding proteins suggests that there was a low abundance of members of the green algal lineage in the sample, which would be consistent with phylogenetic affiliation of those rRNA reads obtained within the metatranscriptome

(section 7.4.2) but not consistent with the plastid 16S rRNA clone libraries in chapter 4.



**Figure 7.10. Proposed pathway for C<sub>4</sub> photosynthesis in marine diatoms (Parker *et al.*, 2008). See text for description of the mechanism. Enzyme abbreviations: PEPCase: phosphoenolpyruvate carboxylase; PPDK: pyruvate-phosphate dikinase; PYC: pyruvate carboxylase; CA: carbonic anhydrase; RUBISCO: ribulose-1,5-bisphosphate carboxylase/oxygenase; MDH: malate dehydrogenase; ME: malic enzyme; PEPCK: phosphoenolpyruvate carboxykinase**

Only a single transcript encoding a potential RuBisCO was found in the metatranscriptome. The encoded ORF was phylogenetically most closely related to a Dinophyceae RuBisCO. The affinity of Dinophyceae RuBisCO for CO<sub>2</sub> is lower than that of RuBisCO's found in red and green algae, coccolithophores or diatoms (Tortell, 2000). However, the cellular C concentration factor (the concentration of intracellular inorganic C relative to external inorganic C) is greater in members of

the Dinophyceae than in any of the other lineages, suggesting that dinophytes are able to compensate for their lower RuBisCO specificity factor by having a more effective CCM (Tortell, 2000).

#### **7.4.5. Metatranscriptomic insights into nutrient acquisition and assimilation**

Transcripts relating to a wide range of transporters involved in nutrient acquisition are found in the metatranscriptome. This includes transporters for inorganic (nitrate and ammonium) and organic (such as urea and amino acids) forms of nitrogen as well as transporters for other essential nutrients such as sulphate. Transcripts relating to nitrate reductase, nitrite reductase, glutamate synthase were also present in the metatranscriptome as well as evidence for the ability of organisms present in the sample to be able to catabolise amino acids with the presences of transaminases. The ability to utilise various forms of nitrogen is similar to that found in the diatom, *T. pseudonana* (Armbrust *et al.*, 2004).

Iron is one of the most effective redox elements on Earth. As a result it is an essential cofactor for many enzymes as well as components of the photosynthetic apparatus. However, marine surface waters are often iron deplete, requiring input from aeolian dust and/or periodic inputs from iron rich deeper waters to sustain microbial growth (Cassar *et al.*, 2007). Hence, it is not surprising that the metatranscriptome included transcripts encoding components of a high affinity iron transport system i.e. ferric reductase and an ATP dependent ABC transporter. The acquisition of iron in marine phytoplankton is still relatively poorly understood. In the genome of *Ostreococcus* no components of a typical eukaryote-like high affinity iron transport system were identified (Palenik *et al.*, 2007). In contrast, the diatom *T. pseudonana* appears to encode a system which suggests a ferroxidase/permease pathway (Kustka *et al.*,

2007) whilst in *P. tricornutum* a ferric reductase and a putative ferrichrome binding protein indicative of bacterial siderophore mediated systems is present (Allen *et al.*, 2008). The presence of transcripts encoding ferric reductase in the metatranscriptome on top of reads encoding components of an ABC transporter system, and hence perhaps a component of a bacterial siderophore based system, as well as a eukaryotic permease system suggests the potential for both systems in members of this community.

Interestingly, the only transcripts encoding electron carriers between the cytochrome  $b_6/f$  complex and photosystem I within the metatranscriptome were those encoding cytochrome  $c_6$ . This is intriguing, and contrasts with the data obtained from the cultured clade VIIA prasinophyte and *Ochromonas* cDNA libraries (see chapter 6), since cytochrome  $c_6$  is an iron requiring electron carrier, that is generally found to be replaced by the copper-requiring plastocyanin under iron deficient conditions as has been found in the genomes of the prasinophytes *O. tauri* and *O. lucimarinus* and the open ocean diatom *T. oceanica* (Peers and Price, 2006; Palenik *et al.*, 2007). Many plastocyanin-containing phytoplankton have been shown to be able to synthesise cytochrome  $c_6$  under copper deficient conditions (Peers and Price, 2006). The absence of reads relating to the copper requiring plastocyanin in the metatranscriptome suggests that the environmental conditions in the southern gyre are such that they favour the use of the iron-requiring cytochrome  $c_6$  instead of the copper-requiring plastocyanin. This is surprising as the south Atlantic gyre region is regarded as having lower levels of iron than the northern hemisphere regions due to a lack of atmospheric dust inputs (Bowie *et al.*, 2002). The dominance of diatoms in

the metatranscriptome could account for the lack of plastocyanin transcripts and many diatom species are believed to only contain cytochrome  $c_6$ .

Transcripts encoding both ferredoxin and flavodoxin were present in the metatranscriptome. Ferredoxin is an iron-requiring electron carrier whilst flavodoxin does not require iron. In *P. tricornutum* the gene encoding flavodoxin has been shown to be significantly upregulated during iron deficiency (Allen *et al.*, 2008), presumably replacing ferredoxin as the electron carrier under these conditions, and thus helping to reduce the cellular demands for iron. The presence of transcripts encoding both of these photosynthetic electron transfer chain components in the metatranscriptome, suggests that some members of the PPE population in this southern gyre community may be experiencing iron stress, whilst others are not, the latter reinforced by the presence of transcripts encoding cytochrome  $c_6$  (see above).

## **7.5 Conclusions and further work**

This study undertook a targeted approach to study gene expression in a natural PPE community in the south Atlantic gyre. The protocol developed in this study, commencing with flow sorting PPE cells for subsequent RNA extraction, mRNA enrichment and amplification appears appropriate for this purpose. However, further work should address any potential biases in this amplification process, perhaps using a cultured PPE so that this pipeline can be directly compared to that used for cDNA library construction (see chapter 6).

The southern gyre metatranscriptome was dominated by reads that affiliated with the classes Dinophyceae and Bacillariophyceae. However, many of the top hits against the BLASTx database indicated that a large proportion of the reads could not be

phylogenetically assigned. Of those that were assigned the majority were related to Bacillariophyceae. This may reflect a bias in the GenBank database towards those classes/organisms for which complete genome sequences are available. In the future, as more PPE genomes are completed, knowledge of the genes present within various PPE lineages will undoubtedly increase, a fact that will significantly aid assignment of environmental transcripts. As evidenced in this chapter, only a small percentage of the predicted ORFs were able to be annotated against public databases. That said, the metatranscriptomic library constructed here was a first step towards understanding the functional capabilities of natural PPE communities. It demonstrated the PPE community in this south Atlantic gyre station had the capacity for a C<sub>4</sub> type photosynthesis/CCM system as well giving an interesting insight into the potential environmental factors that these organisms are responding to. Indeed, the presence of transcripts encoding both cytochrome c<sub>6</sub> and ferredoxin potentially suggests the community is iron-sufficient in this environment. Further studies in different oceanic regions would allow for functional metabolic and nutrient acquisition comparisons between environments differing in their nutrient regimes. Certainly, a greater understanding of the function of PPEs, that are integral to marine carbon cycling, will be obtained using a combination of the phylogenetic studies (chapter 4), culture studies (chapter 5 and 6) and transcriptomic studies (chapter 7) used here. Future studies should develop novel strategies for culturing these PPEs (as attempted in chapter 5) or use culture-independent e.g. single cell genomic approaches (Hutchison and Venter, 2006), to reveal further insights into the functional potential of these organisms.

## **Chapter 8**

### **General discussion and future directions**



## 8.1 General discussion

An initial aim of this study was to obtain taxonomic information about PPE populations flow sorted from the SOLAS cruise in 2008 (Chapter 3) and for which information on the CO<sub>2</sub> fixation rates of these groups had been obtained. This showed that pico-prymnesiophytes, mostly belonging to lineages with no close cultured counterpart, were responsible for up to 38% of the total primary production in the sub-tropical and tropical north-east Atlantic Ocean (see Jardillier *et al.*, 2010).

A subsequent major aim of this PhD was to assess the taxonomic composition and distribution of specific PPE groups along an Atlantic Meridional Transect (AMT18) i.e. at the ocean basin-scale using molecular ecological approaches encompassing a range of PCR primer sets. To meet this goal DNA was extracted from size fractionated samples as well as flow cytometrically sorted PPEs and both nuclear 18S rRNA and plastid biased 16S rRNA genes were used to identify dominant phylotypes and map their distribution, correlating these patterns with measured environmental variables (Chapter 4).

In the total DNA filtered samples, Prymnesiophyceae were seen to be the dominant class throughout the transect, and at all depths, when using the marine-algal plastid biased 16S rRNA gene primers. Chrysophyceae also accounted for a significant proportion of the clones in these libraries, but their contribution decreased with depth, a feature similar to the observations of Kirkham *et al.*, (2011) on AMT 15.

In the flow cytometrically sorted cells Prymnesiophyceae and Chrysophyceae showed a complimentary distribution at the 55% light level. Prymnesiophyceae were the dominant component in the clone libraries in the temperate regions (stations 12

and 91) whilst Chrysophyceae were more prominent in the clone libraries in the tropical and equatorial regions (stations 25, 56, 70). Such a feature was also observed along AMT15 in bulk PPE populations (Kirkham *et al.*, 2011) suggesting this might be an annual feature of these Atlantic waters. Prymnesiophyceae OTUs were phylogenetically related largely to members of *Chrysochromulina* clade B2 with smaller contributions from *Chrysochromulina* clade B1, Coccolithales and Prym 16S-II. At the DCM, clone libraries had a different composition with Prasinophyceae being more prominent and the fraction accounted for by Prymnesiophyceae being diminished. The Prasinophyceae component largely comprised OTUs clustering with clade Pras 16S-VIII and Prasinococcales.

At the 55% light level a latitudinal diversity gradient was observed with a peak in diversity being observed near the equator although the trend was not statistically significant. A similar diversity gradient has been observed in bacterioplankton (see Fuhrman *et al.*, 2008) but analysis of an increased number of stations would be required to have a statistically significant trend in this study with suggestions that the higher temperatures at lower latitudes increase the kinetics of biological processes (Fuhrman *et al.*, 2008). At the DCM the equatorial region showed the lowest diversity suggesting that different factors were affecting diversity at the DCM with the possibility that nutrient concentrations were a greater driving force at the DCM than latitude and temperature.

The nuclear 18S rRNA gene clone libraries were dominated by OTUs phylogenetically related to Prasinophyceae and Dinophyceae. The Prasinophyceae component was comprised of OTUs closely related to prasinophyte clade VIIA and VIIB which have been previously observed in the Pacific Ocean (Guillou *et al.*,

2004). A proportion of the clones in these libraries were accounted for by OTUs related to the heterotrophic lineage Syndiniales, although the use of flow cytometric sorting appears to have reduced the bias towards heterotrophic organisms often observed in nuclear 18S rRNA gene clone libraries (Vaulot *et al.*, 2002).

The clone library approach allows for a finer scale of phylogenetic discrimination than other approaches such as FISH and HPLC. However, there are limitations to the clone library approach, such as PCR bias and cloning biases. Different cloning vectors have been shown to give different patterns of community structure (Palatinszky *et al.*, 2011). However, the results obtained here are in broad agreement with previous studies (e.g. Kirkham *et al.*, 2011). To avoid possible cloning biases approaches such as direct pyrosequencing of rRNA amplicons or single cell genomics could be undertaken.

Attempts were made to obtain into culture novel PPE species/genera (Chapter 5) which were representative of the environmental sequences contained within the constructed clone libraries. However, despite various different culture methods being employed the majority of isolates were related to *Chlorella* (Trebouxiophyceae) and *Prasinoderma* (Prasinococcales), organisms which have been typically found in other PPE culture isolation attempts (see Le Gall *et al.*, 2008). Some of the isolates though were more typical of the environmental sequences found in the 18S rRNA gene clone libraries including a clade VIIA prasinophyte, as well as a *Pelagomonas* sp. isolate. Unfortunately, isolates were not obtained typical of the Prymnesiophyceae and Chrysophyceae species which dominated the oligotrophic open ocean regions and further efforts would be required in order to isolate these.

To gain more information into the functional capability of PPE groups found to be important along AMT18, a transcriptomic (cDNA library) approach was performed using a previously isolated marine *Ochromonas* sp. (CCMP584), as the most appropriate currently available cultured representative of marine Chrysophyceae, and a clade VIIA prasinophyte (RCC1124) (Chapter 6). Phylogenetic analysis showed that the majority of the BLASTx hits in the RCC1124 transcriptome were related to Chlorophyta. However, for the *Ochromonas* sample very few of the reads were related to Chrysophyceae with the majority related to Trebouxiophyceae and the presence of 18S rRNA reads related to *Chlorella* suggesting the culture was not uni-algal. One of the major findings from the RCC1124 transcriptome was the presence of transcripts encoding enzymes involved in C<sub>4</sub> photosynthesis with a proposed pathway similar to that postulated for *Ostreococcus* (Derelle *et al.*, 2006). Both transcriptomes contained transcripts encoding enzymes and transporters for the acquisition and utilisation of both oxidised and reduced forms of nitrogen, but the *Ochromonas* transcriptome also contained transcripts encoding the majority of the enzymes required for a functioning urea cycle, as has previously been observed in marine diatoms (Armbrust *et al.*, 2004; Bowler *et al.*, 2008). Both transcriptomes contained high affinity Fe acquisition systems with the prasinophyte transcriptome possessing reads encoding a bacterial Fe scavenging siderophore, enterobactin. The ability to transport iron by this means has been previously hypothesised for both diatoms and *Ostreococcus lucimarinus* (Armbrust *et al.*, 2004; Palenik *et al.*, 2007).

To begin to transcend the boundary between culture studies and the environment, a pipeline was established for analysing the metatranscriptome of a natural PPE population so as to begin to define the environmental variables these organisms are

responding to *in situ* (Chapter 7). This entailed the extraction of total RNA and enrichment of mRNA from flow cytometrically sorted PPE cells from a station in the southern gyre along AMT18. The use of terminator<sup>TM</sup> exonuclease appeared to be the method of choice for the enrichment of mRNA from the small amount of starting material available since only 0.62% of the resulting contigs contained rRNA reads. Phylogenetic affiliation of these reads was mainly with the classes Dinophyceae and Bacillariophyceae. This is in contrast to the marine-algal plastid biased 16S rRNA library from the same site where Prymnesiophyceae and Trebouxiophyceae were the dominant classes. The dominance of Bacillariophyceae reads in the top BLASTx hits may in part reflect the presence of sequenced genomes of this class in databases. This may also explain the paucity of Chrysophyceae hits, since there is little or no sequence information available for these organisms, a fact not helped with the potential non uni-algal nature of the *Ochromonas* culture used for cDNA library construction (Chapter 6). Functional classification of the environmental transcriptome reads showed that of those reads which could be assigned, a large proportion could only be assigned to a general function. The KOG database, with only seven eukaryote genomes represented, restricts the ability to assign functions to reads of unrepresented organisms. The addition of further phytoplankton genomes to this database is a real prerequisite for further study, and would certainly increase the predictive power of this transcriptomic approach. Even so, a major finding of the metatranscriptomic analysis was the presence of contigs putatively encoding enzymes involved in the C<sub>4</sub> photosynthesis pathway similar to that proposed for diatoms (Reinfelder *et al.*, 2004).

## 8.2 Future directions

Ecological, biochemical and genetic studies of PPEs are clearly lagging behind those of prokaryotic phytoplankton. That the taxonomic composition of PPEs is still poorly understood is evidenced by the very recent discovery of novel groups such as picobiliphytes (Not *et al.*, 2007) and rappemonads (Kim *et al.*, 2010), and the plethora of novel sequences obtained here. Further characterisation of the PPE component, utilising flow cytometric sorting and a variety of primer sets, particularly primers targeting functional genes, would enable an even greater understanding of their taxonomic affiliation, particularly at a fine scale resolution i.e. beyond the use of rRNA genes. Increased surveys especially in open ocean regions will enable a better understanding of the distribution of PPEs and allow for improved correlations to be made with nutrient concentrations and other ancillary measurements, so as to begin to determine the important factors controlling the abundance and distribution of specific PPE groups. In turn, this will help us greatly to understand controls on global marine CO<sub>2</sub> fixation. Certainly, increased knowledge of the species present will also enable the design of new primer sets (see above) which would potentially allow taxonomic discrimination at the level of ecotypes allowing much more careful correlation of genotype abundances as a function of environmental parameters.

Undoubtedly, further attempts to obtain into culture new PPE isolates, especially from oligotrophic regions will be required to increase our basic knowledge of these organisms. The application of single cell sorting and the use of dilute culture medium will be a first step towards this. Recently, novel methods have been developed to isolate marine bacterioplankton (see for example Alain and Querellou, 2009; Bruns *et al.*, 2002) and in the future some of these methods could be adapted

for the isolation of PPEs. For example, attempts to culture marine bacterioplankton are highlighting the importance of cell to cell interactions in marine systems. Methods such as single cell encapsulation in a continuous flow of medium allows for the growth of clonal colonies whilst metabolites and signalling molecules are able to freely pass through the culture (Zengler *et al.*, 2002). Analysis of mixed species cultures would help to identify specific metabolites or micronutrients which are vital for the growth of a specific organism. Identification of these metabolites would then allow for the successful isolation of these organisms into monoculture (Alain and Querellou, 2009). For instance, the use of cAMP has enabled the cultivation of novel bacterioplankton which were previously in low abundance in the community (Bruns *et al.*, 2002). The ability to isolate into culture novel PPEs which are environmentally important, such as prymnesiophytes and chrysophytes, will allow physiological and genetic studies of these organisms increasing our knowledge of their metabolic capabilities and potentially substantially increasing our understanding of the biology of arguably the most important components of the marine PPE community.

So far, only five species classified as PPEs have had their genomes sequenced. Of these, four belong to the Mamiellophyceae (Derelle *et al.*, 2006; Palenik *et al.*, 2007; Worden *et al.*, 2009). However, Mamiellophyceae, although a significant component of the PPE community in coastal regions, are much less important in open ocean regions (see Not *et al.*, 2004, Not *et al.*, 2008; this work). Future directions for genomic studies of PPEs require the sequencing of genomes from organisms which are more typical of open ocean areas such as those belonging to the Prymnesiophyceae and Chrysophyceae. However, with the lack of representative

cultures (see above) a metagenomic/transcriptomic approach will be one way which could reveal some of the genomic secrets of these unculturable organisms. There are however, limitations to these approaches. The resolution of metagenomic/transcriptomic libraries is generally quite low and only the dominant populations produce sufficient coverage to enable sequence assembly (Warnecke and Hugenholtz, 2007). One way in which to alleviate this issue is to target a subsample of the environment, such as the specific targeting of PPEs, via flow cytometry, in this study.

A second limitation is in identifying the source species of specific DNA sequences. Metagenomic sequencing of simple environments has led to the ability to reconstruct genomes of the predominant members. However, even with large sequencing efforts, genome assembly and metabolic reconstruction of members of complex environments has proved problematic (Riesenfeld *et al.*, 2004). The isolation of single cells from complex environments, by methods such as flow cytometry or FISH, can form the basis of single cell genomics (Ishoei *et al.*, 2008). Amplification of the DNA from single cells by multiple displacement amplification (MDA) gives sufficient quantities for subsequent analysis. Single cell genomics allows the linkage of metabolic functions to specific lineages (Stepanauskas and Sieracki, 2007). In metagenomic samples, unless the metabolic gene is located close to a phylogenetic marker (e.g. rRNA) then taxonomic assignment is difficult.

Furthermore, single cell genomics can circumvent some of the biases associated with PCR and cloning based taxonomic studies. Heywood *et al.*, (2011) investigated the composition of marine heterotrophic protists using both conventional clone library and single cell genomic approaches. These two methods indicated a different



composition of the environmental sample possibly due to biases, such as rRNA gene number, in the clone library approach. Thus, single cell genomic methods have the potential to avoid PCR biases and also enable the linkage of phylogeny with metabolic potential

The third limitation is that the libraries can only provide information on fragments which have homology to known proteins (Warnecke and Hugenholtz, 2007).

Clearly, an integration of taxonomic, culture isolation and genomic studies will be needed in order to fully understand the function of PPEs in the marine environment with knowledge from one avenue used to improve the outcomes in others.

## Bibliography

**Agusti, S. & Duarte, C. M.** (1999) Phytoplankton chlorophyll *a* distribution and water column stability in the central Atlantic Ocean. *Oceanologica Acta*, **22**, 193-203.

**Aiken, J. & Bale, A. J.** (2000) An introduction to the Atlantic Meridional Transect (AMT) programme. *Progress in Oceanography*, **45**, 251-256.

**Aiken, J., Pradhan, Y., Barlow, R., Lavender, S., Poulton, A., Holligan, P. & Hardman-Mountford, N.** (2009) Phytoplankton pigments and functional types in the Atlantic Ocean: A decadal assessment, 1995-2005. *Deep-Sea Research Part II-Topical Studies in Oceanography*, **56**, 899-917.

**Alain, K. & Querellou, J.** (2009) Cultivating the uncultured: limits, advances and future challenges. *Extremophiles*, **13**, 583-594.

**Allen, A. E., Vardi, A. & Bowler, C.** (2006) An ecological and evolutionary context for integrated nitrogen metabolism and related signalling pathways in marine diatoms. *Current Opinion in Plant Biology*, **9**, 264-273.

**Allen, A. E., LaRoche, J., Maheswari, U., Lommer, M., Schauer, N., Lopez, P. J., Finazzi, G., Fernie, A. R. & Bowler, C.** (2008) Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 10438-10443.

**Allen, A. E., Dupont, C. L., Obornik, M., Horak, A., Nunes-Nesi, A., McCrow, J. P., Zheng, H., Johnson, D. A., Hu, H., Fernie, A. R. & Bowler, C.** (2011) Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature*, **473**, 203-209.

**Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J.** (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403-410.

**Andersen, R.A., Bidigare, R.R., Keller, M.D., & Latasa, M.** (1999) A comparison of HPLC pigment signatures and electron microscopic observations for oligotrophic waters of the North Atlantic and Pacific Oceans. *Deep Sea Research II* **43**, 517-537

**Andersen, R.A.** (2005) Algal culturing techniques. Academic Press 578pp.

**Andersen, R. A., Saunders, G. W., Paskind, M. P. & Sexton, J. P.** (1993) Ultrastructure and 18S ribosomal RNA gene sequence for *Pelagomonas calceolata* gen. et sp. nov. and the description of a new algal class, the Pelagophyceae classis nov. *Journal of Phycology*, **29**, 701-715.

**Andersen, R. A., Bidigare, R. R., Keller, M. D. & Latasa, M.** (1996) A comparison of HPLC pigment signatures and electron microscopic observations for oligotrophic waters of the North Atlantic and Pacific Oceans. *Deep-Sea Research Part II-Topical Studies in Oceanography*, **43**, 517-537.

**Archibald, J. M.** (2006) Genome complexity in a lean, mean photosynthetic machine. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 11433-11434.

**Archibald, J. M.** (2009a) The puzzle of plastid evolution. *Current Biology*, **19**, R81-R88.

**Archibald, J. M.** (2009b) Green evolution, green revolution. *Science*, **324**, 191-192.

**Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., Zhou, S. G., Allen, A. E., Apt, K. E., Bechner, M., Brzezinski, M. A., Chaal, B. K., Chiovitti, A., Davis, A. K., Demarest, M. S., Detter, J. C., Glavina, T., Goodstein, D., Hadi, M. Z., Hellsten, U., Hildebrand, M., Jenkins, B. D., Jurka, J., Kapitonov, V. V., Kroger, N., Lau, W. W. Y., Lane, T. W., Larimer, F. W., Lippmeier, J. C., Lucas, S., Medina, M., Montsant, A., Obornik, M., Parker, M. S., Palenik, B., Pazour, G. J., Richardson, P. M., Rynearson, T. A., Saito, M. A., Schwartz, D. C., Thamatrakoln, K., Valentin, K., Vardi, A., Wilkerson, F. P. & Rokhsar, D. S.** (2004) The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. *Science*, **306**, 79-86.

**Azam, F.** (1998) Microbial control of oceanic carbon flux: The plot thickens. *Science*, **280**, 694-696.

**Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyerreil, L. A. & Thingstad, F.** (1983) The ecological role of water column microbes in the sea. *Marine Ecology-Progress Series*, **10**, 257-263.

**Barber, J.** (2008) Photosynthetic generation of oxygen. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **363**, 2665-2674.

**Barton, A. D., Dutkiewicz, S., Flierl, G., Bragg, J. & Follows, M. J.** (2010) Patterns of diversity in marine phytoplankton. *Science*, **327**, 1509-1511.

**Baudoux, A. C., Veldhuis, M. J. W., Noordeloos, A. A. M., van Noort, G. & Brussaard, C. P. D.** (2008) Estimates of virus- vs. grazing induced mortality of picophytoplankton in the North Sea during summer. *Aquatic Microbial Ecology*, **52**, 69-82.

- Bec, B., Collos, Y., Vaquer, A., Mouillot, D. & Souchu, P.** (2008) Growth rate peaks at intermediate cell size in marine photosynthetic picoeukaryotes. *Limnology and Oceanography*, **53**, 863-867.
- Behrenfeld, M. J. & Kolber, Z. S.** (1999) Widespread iron limitation of phytoplankton in the South Pacific Ocean. *Science*, **283**, 840-843.
- Berney, C., Fahrni, J. & Pawlowski, J.** (2004) How many novel eukaryotic 'kingdoms'? Pitfalls and limitations of environmental DNA surveys. *BMC Biology*, **2**, 13.
- Bhattacharya, D., Yoon, H. S. & Hackett, J. D.** (2004) Photosynthetic eukaryotes unite: endosymbiosis connects the dots. *Bioessays*, **26**, 50-60.
- Biegala, I. C., Not, F., Vaultot, D. & Simon, N.** (2003) Quantitative assessment of picoeukaryotes in the natural environment by using taxon-specific oligonucleotide probes in association with tyramide signal amplification-fluorescence in situ hybridization and flow cytometry. *Applied and Environmental Microbiology*, **69**, 5519-5529.
- Blanc, G., Duncan, G., Agarkova, I., Borodovsky, M., Gurnon, J., Kuo, A., Lindquist, E., Lucas, S., Pangilinan, J., Polle, J., Salamov, A., Terry, A., Yamada, T., Dunigan, D. D., Grigoriev, I. V., Claverie, J.-M. & Van Etten, J. L.** (2010) The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell*, **22**, 2943-2955.
- Bonente, G., Howes, B. D., Caffarri, S., Smulevich, G. & Bassi, R.** (2008) Interactions between the photosystem II subunit PsbS and xanthophylls studied in vivo and in vitro. *Journal of Biological Chemistry*, **283**, 8434-8445.
- Borcard, D., Legendre, P. & Drapeau, P.** (1992) Partialling out the spatial component of ecological variation. *Ecology*, **73**, 1045-1055.
- Bowie, A. R., Whitworth, D. J., Achterberg, E. P., Mantoura, R. F. C. & Worsfold, P. J.** (2002) Biogeochemistry of Fe and other trace elements (Al, Co, Ni) in the upper Atlantic Ocean. *Deep-Sea Research Part I-Oceanographic Research Papers*, **49**, 605-636.
- Bowler, C., Vanmontagu, M. & Inze, D.** (1992) Superoxide dismutase and stress tolerance. *Annual Review of Plant Physiology and Plant Molecular Biology*, **43**, 83-116.

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otilar, R. P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Verret, F., Berges, J. A., Brownlee, C., Cadoret, J. P., Chiovitti, A., Choi, C. J., Coesel, S., De Martino, A., Detter, J. C., Durkin, C., Falciatore, A., Fournet, J., Haruta, M., Huysman, M. J. J., Jenkins, B. D., Jiroutova, K., Jorgensen, R. E., Joubert, Y., Kaplan, A., Kroger, N., Kroth, P. G., La Roche, J., Lindquist, E., Lommer, M., Martin-Jezequel, V., Lopez, P. J., Lucas, S., Mangogna, M., McGinnis, K., Medlin, L. K., Montsant, A., Oudot-Le Secq, M. P., Napoli, C., Obornik, M., Parker, M. S., Petit, J. L., Porcel, B. M., Poulsen, N., Robison, M., Rychlewski, L., Rynearson, T. A., Schmutz, J., Shapiro, H., Siaut, M., Stanley, M., Sussman, M. R., Taylor, A. R., Vardi, A., von Dassow, P., Vyverman, W., Willis, A., Wyrwicz, L. S., Rokhsar, D. S., Weissenbach, J., Armbrust, E. V., Green, B. R., Van De Peer, Y. & Grigoriev, I. V. (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature*, **456**, 239-244.

Boyd, P. W., Jickells, T., Law, C. S., Blain, S., Boyle, E. A., Buesseler, K. O., Coale, K. H., Cullen, J. J., de Baar, H. J. W., Follows, M., Harvey, M., Lancelot, C., Levasseur, M., Owens, N. P. J., Pollard, R., Rivkin, R. B., Sarmiento, J., Schoemann, V., Smetacek, V., Takeda, S., Tsuda, A., Turner, S. & Watson, A. J. (2007) Mesoscale iron enrichment experiments 1993-2005: Synthesis and future directions. *Science*, **315**, 612-617.

Bravo-Sierra, E. & Hernandez-Becerril, D. U. (2003) Parmales (Chrysophyceae) from the Gulf of Tehuantepec, Mexico, including the description of a new species, *Tetraparma insecta* sp. nov., and a proposal to the taxonomy of the group. *Journal of Phycology*, **39**, 577-583.

Brunet, C., Casotti, R., Vantrepotte, V., Corato, F. & Conversano, F. (2006) Picophytoplankton diversity and photoacclimation in the Strait of Sicily (Mediterranean Sea) in summer. I. Mesoscale variations. *Aquatic Microbial Ecology*, **44**, 127-141.

Bruns, A., Cypionka, H. & Overmann, J. (2002) Cyclic AMP and acyl homoserine lactones increase the cultivation efficiency of heterotrophic bacteria from the central Baltic Sea. *Applied and Environmental Microbiology*, **68**, 3978-3987.

Cashmore, A. R., Jarillo, J. A., Wu, Y. J. & Liu, D. M. (1999) Cryptochromes: Blue light receptors for plants and animals. *Science*, **284**, 760-765.

Cassar, N., Bender, M. L., Barnett, B. A., Fan, S., Moxim, W. J., Levy, H., II & Tilbrook, B. (2007) The Southern Ocean biological response to aeolian iron deposition. *Science*, **317**, 1067-1070.

- Cermeno, P., Maranon, E., Harbour, D., Figueiras, F. G., Crespo, B. G., Huete-Ortega, M., Varela, M. & Harris, R. P.** (2008) Resource levels, allometric scaling of population abundance, and marine phytoplankton diversity. *Limnology and Oceanography*, **53**, 312-318.
- Chambouvet, A., Morin, P., Marie, D. & Guillou, L.** (2008) Control of toxic marine dinoflagellate blooms by serial parasitic killers. *Science*, **322**, 1254-1257.
- Chesnick, J. M., Kooistra, W., Wellbrock, U. & Medlin, L. K.** (1997) Ribosomal RNA analysis indicates a benthic pennate diatom ancestry for the endosymbionts of the dinoflagellates *Peridinium foliaceum* and *Peridinium balticum* (Pyrrhophyta). *Journal of Eukaryotic Microbiology*, **44**, 314-320.
- Cheung, M. K., Au, C. H., Chu, K. H., Kwan, H. S. & Wong, C. K.** (2010) Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. *ISME Journal*, **4**, 1053-1059.
- Chistoserdova, L.** (2010) Functional metagenomics: recent advances and future challenges. *Biotechnology and Genetic Engineering Reviews*, Vol 26, **26**, 335-351.
- Colwell, R. K. & Coddington, J. A.** (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **345**, 101-118.
- Courties, C., Vaquer, A., Troussellier, M., Lautier, J., Chretiennotdinet, M. J., Neveux, J., Machado, C. & Claustre, H.** (1994) Smallest eukaryotic organism. *Nature*, **370**, 255-255.
- Crosbie, N. D., Pockl, M. & Weisse, T.** (2003) Rapid establishment of clonal isolates of freshwater autotrophic picoplankton by single-cell and single-colony sorting. *Journal of Microbiological Methods*, **55**, 361-370.
- Crow, G. E.** (1993) Species diversity in aquatic angiosperms: Latitudinal patterns. *Aquatic Botany*, **44**, 229-258.
- Cullen, J. J., Lewis, M. R., Davis, C. O. & Barber, R. T.** (1992) Photosynthetic characteristics and estimated growth rates indicate grazing is the proximate control of primary production in the equatorial Pacific. *Journal of Geophysical Research-Oceans*, **97**, 639-654.
- Cutter, G. A. & Cutter, L. S.** (2004) Selenium biogeochemistry in the San Francisco Bay estuary: changes in water column behavior. *Estuarine Coastal and Shelf Science*, **61**, 463-476.

**Cuvelier, M. L., Allen, A. E., Monier, A., McCrow, J. P., Messie, M., Tringe, S. G., Woyke, T., Welsh, R. M., Ishoey, T., Lee, J. H., Binder, B. J., DuPont, C. L., Latasa, M., Guigand, C., Buck, K. R., Hilton, J., Thiagarajan, M., Caler, E., Read, B., Lasken, R. S., Chavez, F. P. & Worden, A. Z.** (2010) Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 14679-14684.

**Davey, M., Tarran, G. A., Mills, M. M., Ridame, C., Geider, R. J. & LaRoche, J.** (2008) Nutrient limitation of picophytoplankton photosynthesis and growth in the tropical North Atlantic. *Limnology and Oceanography*, **53**, 1722-1733.

**DeLong, E. F. & Karl, D. M.** (2005) Genomic perspectives in microbial oceanography. *Nature*, **437**, 336-342.

**Derelle, E., Ferraz, C., Rombauts, S., Rouze, P., Worden, A. Z., Robbens, S., Partensky, F., Degroeve, S., Echeynie, S., Cooke, R., Saeys, Y., Wuyts, J., Jabbari, K., Bowler, C., Panaud, O., Piegu, B., Ball, S. G., Ral, J. P., Bouget, F. Y., Piganeau, G., De Baets, B., Picard, A., Delseny, M., Demaille, J., Van de Peer, Y. & Moreau, H.** (2006) Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 11647-11652.

**Deutschbauer, A. M., Chivian, D. & Arkin, A. P.** (2006) Genomics for environmental microbiology. *Current Opinion in Biotechnology*, **17**, 229-235.

**Devred, E., Sathyendranath, S. & Platt, T.** (2007) Delineation of ecological provinces using ocean colour radiometry. *Marine Ecology-Progress Series*, **346**, 1-13.

**Diez, B., Pedros-Alio, C. & Massana, R.** (2001) Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Applied and Environmental Microbiology*, **67**, 2932-2941.

**Douglas, S., Zauner, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L. T., Wu, X. N., Reith, M., Cavalier-Smith, T. & Maier, U. G.** (2001) The highly reduced genome of an enslaved algal nucleus. *Nature*, **410**, 1091-1096.

**Ducklow, H. W., Steinberg, D. K. & Buesseler, K. O.** (2001) Upper ocean carbon export and the biological pump. *Oceanography*, **14**, 50-58.

- Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I. M., Barbe, V., Duprat, S., Galperin, M. Y., Koonin, E. V., Le Gall, F., Makarova, K. S., Ostrowski, M., Oztas, S., Robert, C., Rogozin, I. B., Scanlan, D. J., de Marsac, N. T., Weissenbach, J., Wincker, P., Wolf, Y. I. & Hess, W. R. (2003) Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 10020-10025.
- Dufresne, A., Ostrowski, M., Scanlan, D. J., Garczarek, L., Mazard, S., Palenik, B. P., Paulsen, I. T., de Marsac, N. T., Wincker, P., Dossat, C., Ferriera, S., Johnson, J., Post, A. F., Hess, W. R. & Partensky, F. (2008) Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biology*, **9**.
- Dupont, C. L., Buck, K. N., Palenik, B. & Barbeau, K. (2010) Nickel utilization in phytoplankton assemblages from contrasting oceanic regimes. *Deep-Sea Research Part I-Oceanographic Research Papers*, **57**, 553-566.
- Eddy, S. R. (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755-763.
- Emanuelsson, O., Brunak, S., Von Heijne, G. & Nielsen, H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols*, **2**, 953-971.
- Falkowski, P. G. & Godfrey, L. V. (2008) Electrons, life and the evolution of Earth's oxygen cycle. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **363**, 2705-2716.
- Falkowski, P. G. & Oliver, M. J. (2007) Mix and match: how climate selects phytoplankton. *Nature Reviews Microbiology*, **5**, 813-819.
- Falkowski, P. G., Barber, R. T. & Smetacek, V. (1998) Biogeochemical controls and feedbacks on ocean primary production. *Science*, **281**, 200-206.
- Falkowski, P., Scholes, R. J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Hogberg, P., Linder, S., Mackenzie, F. T., Moore, B., Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V. & Steffen, W. (2000) The global carbon cycle: A test of our knowledge of earth as a system. *Science*, **290**, 291-296.
- Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. (1998) Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science*, **281**, 237-240.
- Finn, R. D., Tate, J., Mistry, J., Coghill, P. C., Sammut, S. J., Hotz, H.-R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L. L. & Bateman, A. (2008) The Pfam protein families database. *Nucleic Acids Research*, **36**, D281-D288.



**Foster, R. A., Kuypers, M. M. M., Vagner, T., Paerl, R. W., Musat, N. & Zehr, J. P.** (2011) Nitrogen fixation and transfer in open ocean diatom-cyanobacterial symbioses. *Isme Journal*, **5**, 1484-1493.

**Frias-Lopez, J., Shi, Y., Tyson, G. W., Coleman, M. L., Schuster, S. C., Chisholm, S. W. & DeLong, E. F.** (2008) Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 3805-3810.

**Fuhrman, J. A.** (1999) Marine viruses and their biogeochemical and ecological effects. *Nature*, **399**, 541-548.

**Fuhrman, J. A., Steele, J. A., Hewson, I., Schwalbach, M. S., Brown, M. V., Green, J. L. & Brown, J. H.** (2008) A latitudinal diversity gradient in planktonic marine bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 7774-7778.

**Fuller, N. J., Marie, D., Partensky, F., Vaultot, D., Post, A. F. & Scanlan, D. J.** (2003) Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Applied and Environmental Microbiology*, **69**, 2430-2443.

**Fuller, N. J., Campbell, C., Allen, D. J., Pitt, F. D., Zwirgmaierl, K., Le Gall, F., Vaultot, D. & Scanlan, D. J.** (2006a) Analysis of photosynthetic picoeukaryote diversity at open ocean sites in the Arabian Sea using a PCR biased towards marine algal plastids. *Aquatic Microbial Ecology*, **43**, 79-93.

**Fuller, N. J., Tarran, G. A., Cummings, D. G., Woodward, E. M. S., Orcutt, K. M., Yallop, M., Le Gall, F. & Scanlan, D. J.** (2006b) Molecular analysis of photosynthetic picoeukaryote community structure along an Arabian Sea transect. *Limnology and Oceanography*, **51**, 2502-2514.

**Gadd, G. M. & Raven, J. A.** (2010) Geomicrobiology of eukaryotic microorganisms. *Geomicrobiology Journal*, **27**, 491-519.

**Gast, R. J., Moran, D. M., Dennett, M. R. & Caron, D. A.** (2007) Kleptoplasty in an Antarctic dinoflagellate: caught in evolutionary transition? *Environmental Microbiology*, **9**, 39-45.

**Gescher, C., Metfies, K., Frickenhaus, S., Knefelkamp, B., Wiltshire, K. H. & Medlin, L. K.** (2008) Feasibility of assessing the community composition of prasinophytes at the Helgoland roads sampling site with a DNA microarray. *Applied and Environmental Microbiology*, **74**, 5305-5316.

**Gianoulis, T. A., Raes, J., Patel, P. V., Bjornson, R., Korbel, J. O., Letunic, I., Yamada, T., Paccanaro, A., Jensen, L. J., Snyder, M., Bork, P. & Gerstein, M. B.** (2009) Quantifying environmental adaptation of metabolic pathways in metagenomics. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 1374-1379.

**Gibb, S. W., Barlow, R. G., Cummings, D. G., Rees, N. W., Trees, C. C., Holligan, P. & Suggett, D.** (2000) Surface phytoplankton pigment distributions in the Atlantic Ocean: an assessment of basin scale variability between 50 degrees N and 50 degrees S. *Progress in Oceanography*, **45**, 339-368.

**Gibbs, P. B. & Biggins, J.** (1989) Regulation of the distribution of excitation energy in *Ochromonas danica*, an organism containing a chlorophyll a/c/carotenoid light harvesting antenna. *Photosynthesis Research*, **21**, 81-91.

**Gibbs, S. P.** (1978) Chloroplasts of *Euglena* may have evolved from symbiotic green algae. *Canadian Journal of Botany-Revue Canadienne De Botanique*, **56**, 2883-2889.

**Gifford, S. M., Sharma, S., Rinta-Kanto, J. M. & Moran, M. A.** (2011) Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *ISME Journal*, **5**, 461-472.

**Gilbert, B., McDonald, I. R., Finch, R., Stafford, G. P., Nielsen, A. K. & Murrell, J. C.** (2000) Molecular analysis of the pmo (particulate methane monooxygenase) operons from two type II methanotrophs. *Applied and Environmental Microbiology*, **66**, 966-975.

**Gobler, C. J., Berry, D. L., Dyhrman, S. T., Wilhelm, S. W., Salamov, A., Lobanov, A. V., Zhang, Y., Collier, J. L., Wurch, L. L., Kustka, A. B., Dill, B. D., Shah, M., VerBerkmoes, N. C., Kuo, A., Terry, A., Pangilinan, J., Lindquist, E. A., Lucas, S., Paulsen, I. T., Hattenrath-Lehmann, T. K., Talmage, S. C., Walker, E. A., Koch, F., Burson, A. M., Marcoval, M. A., Tang, Y.-Z., LeClerc, G. R., Coyne, K. J., Berg, G. M., Bertrand, E. M., Saito, M. A., Gladyshev, V. N. & Grigoriev, I. V.** (2011) Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 4352-4357.

**Good, I. J.** (1953) The population frequencies of species and the estimation of population parameters. *Biometrika*, **40**, 237-264.

**Grant, S., Grant, W. D., Cowan, D. A., Jones, B. E., Ma, Y. H., Ventosa, A. & Heaphy, S.** (2006) Identification of eukaryotic open reading frames in metagenomic cDNA libraries made from environmental samples. *Applied and Environmental Microbiology*, **72**, 135-143.

- Green, B. R.** (2011) After the primary endosymbiosis: an update on the chromalveolate hypothesis and the origins of algae with Chl *c*. *Photosynthesis Research*, **107**, 103-115.
- Grob, C., Ulloa, O., Claustre, H., Huot, Y., Alarcon, G. & Marie, D.** (2007) Contribution of picoplankton to the total particulate organic carbon concentration in the eastern South Pacific. *Biogeosciences*, **4**, 837-852.
- Guazzaroni, M.-E., Beloqui, A., Golyshin, P. N. & Ferrer, M.** (2009) Metagenomics as a new technological tool to gain scientific knowledge. *World Journal of Microbiology & Biotechnology*, **25**, 945-954.
- Guillard, R. R. & Ryther, J. H.** (1962) Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt, and *Detonula confervacea* (cleve) Gran. *Canadian Journal of Microbiology*, **8**, 229-39.
- Guillou, L., Moon-Van Der Staay, S. Y., Claustre, H., Partensky, F. & Vaulot, D.** (1999) Diversity and abundance of Bolidophyceae (Heterokonta) in two oceanic regions. *Applied and Environmental Microbiology*, **65**, 4528-4536.
- Guillou, L., Eikrem, W., Chretiennot-Dinet, M. J., Le Gall, F., Massana, R., Romari, K., Pedros-Alio, C. & Vaulot, D.** (2004) Diversity of picoplanktonic prasinophytes assessed by direct nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. *Protist*, **155**, 193-214.
- Guillou, L., Viprey, M., Chambouvet, A., Welsh, R. M., Kirkham, A. R., Massana, R., Scanlan, D. J. & Worden, A. Z.** (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environmental Microbiology*, **10**, 3349-3365.
- Hackett, J. D., Yoon, H. S., Bhattacharya, D.** (2006) Genome evolution in dinoflagellates. In *Genome Evolution in the Algae*. Edited by Bhattacharya, D. and Katz, L. Oxford University Press.
- Haft, D. H., Selengut, J. D. & White, O.** (2003) The TIGRFAMs database of protein families. *Nucleic Acids Research*, **31**, 371-373.
- Hall, J. A. & Safi, K.** (2001) The impact of in situ Fe fertilisation on the microbial food web in the Southern Ocean. *Deep-Sea Research Part II-Topical Studies in Oceanography*, **48**, 2591-2613.
- Hamilton, A. K., Lovejoy, C., Galand, P. E. & Ingram, R. G.** (2008) Water masses and biogeography of picoeukaryote assemblages in a cold hydrographically complex system. *Limnology and Oceanography*, **53**, 922-935.

**Handelsman, J.** (2004) Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews*, **68**, 669-+.

**Heywood, J. L., Sieracki, M. E., Bellows, W., Poulton, N. J. & Stepanauskas, R.** (2011) Capturing diversity of marine heterotrophic protists: one cell at a time. *Isme Journal*, **5**, 674-684.

**Hillebrand, H.** (2004) On the generality of the latitudinal diversity gradient. *American Naturalist*, **163**, 192-211.

**Hirata, T., Aiken, J., Hardman-Mountford, N., Smyth, T. J. & Barlow, R. G.** (2008) An absorption model to determine phytoplankton size classes from satellite ocean colour. *Remote Sensing of Environment*, **112**, 3153-3159.

**Howe, C. J., Schlarb-Ridley, B. G., Wastl, J., Purton, S. & Bendall, D. S.** (2006) The novel cytochrome c(6) of chloroplasts: a case of evolutionary bricolage? *Journal of Experimental Botany*, **57**, 13-22.

**Huang, Y., Gilna, P. & Li, W.** (2009) Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics*, **25**, 1338-1340.

**Huber, T., Faulkner, G. & Hugenholtz, P.** (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics*, **20**, 2317-2319.

**Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C.** (2007) MEGAN analysis of metagenomic data. *Genome Research*, **17**, 377-386.

**Hutchins, D. A. & Bruland, K. W.** (1998) Iron-limited diatom growth and Si:N uptake ratios in a coastal upwelling regime. *Nature*, **393**, 561-564.

**Hutchison, C. A. & Venter, J. C.** (2006) Single-cell genomics. *Nature Biotechnology*, **24**, 657-658.

**Ichinomiya, M., Yoshikawa, S., Kamiya, M., Ohki, K., Takaichi, S. & Kuwata, A.** (2011) Isolation and characterization of Parmales (Heterokonta/Heterokonophyta/Stramenopiles) from the Oyashio region, western north Pacific. *Journal of Phycology*, **47**, 144-151.

**Ishoe, T., Woyke, T., Stepanauskas, R., Novotny, M. & Lasken, R. S.** (2008) Genomic sequencing of single microbial cells from environmental samples. *Current Opinion in Microbiology*, **11**, 198-204.

- Jardillier, L., Zubkov, M. V., Pearman, J. & Scanlan, D. J.** (2010) Significant CO<sub>2</sub> fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. *ISME Journal*, **4**, 1180-1192.
- Jickells, T. D., An, Z. S., Andersen, K. K., Baker, A. R., Bergametti, G., Brooks, N., Cao, J. J., Boyd, P. W., Duce, R. A., Hunter, K. A., Kawahata, H., Kubilay, N., laRoche, J., Liss, P. S., Mahowald, N., Prospero, J. M., Ridgwell, A. J., Tegen, I. & Torres, R.** (2005) Global iron connections between desert dust, ocean biogeochemistry, and climate. *Science*, **308**, 67-71.
- John, D. E., Zielinski, B. L. & Paul, J. H.** (2009) Creation of a pilot metatranscriptome library from eukaryotic plankton of a eutrophic bay (Tampa Bay, Florida). *Limnology and Oceanography-Methods*, **7**, 249-259.
- Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M. S. & Chisholm, S. W.** (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science*, **311**, 1737-1740.
- Kan, J., Hanson, T. E., Ginter, J. M., Wang, K. & Chen, F.** (2005) Metaproteomic analysis of Chesapeake Bay microbial communities. *Saline systems*, **1**, 7.
- Kanehisa, M. & Goto, S.** (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, **28**, 27-30.
- Karlsson, J., Clarke, A. K., Chen, Z. Y., Huggins, S. Y., Park, Y. I., Husic, H. D., Moroney, J. V. & Samuelsson, G.** (1998) A novel alpha-type carbonic anhydrase associated with the thylakoid membrane in *Chlamydomonas reinhardtii* is required for growth at ambient CO<sub>2</sub>. *EMBO Journal*, **17**, 1208-1216.
- Kawachi, M., Inouye, I., Honda, D., O'Kelly, C. J., Bailey, J. C., Bidigare, R. R. & Andersen, R. A.** (2002) The Pinguiphyceae classis nov., a new class of photosynthetic stramenopiles whose members produce large amounts of omega-3 fatty acids. *Phycological Research*, **50**, 31-47.
- Keeling, P. J.** (2010) The endosymbiotic origin, diversification and fate of plastids. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **365**, 729-748.
- Keller, M. D., Selvin, R. C., Claus, W. & Guillard, R. R. L.** (1987) Media for the culture of oceanic ultraphytoplankton. *Journal of Phycology*, **23**, 633-638.
- Kettler, G. C., Martiny, A. C., Huang, K., Zucker, J., Coleman, M. L., Rodrigue, S., Chen, F., Lapidus, A., Ferreira, S., Johnson, J., Steglich, C., Church, G. M., Richardson, P. & Chisholm, S. W.** (2007) Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *Plos Genetics*, **3**, 2515-2528.

- Kim, E., Harrison, J. W., Sudek, S., Jones, M. D. M., Wilcox, H. M., Richards, T. A., Worden, A. Z. & Archibald, J. M.** (2010) Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 1496-1500.
- Kirkham, A. R., Jardillier, L. E., Tiganeşcu, A., Pearman, J., Zubkov, M. V. & Scanlan, D. J.** (2011) Basin-scale distribution patterns of photosynthetic picoeukaryotes along an Atlantic Meridional Transect. *Environmental Microbiology*, **13**, 975-990.
- Kondou, Y., Nakazawa, M., Higashi, S., Watanabe, M. & Manabe, K.** (2001) Equal-quantum action spectra indicate fluence-rate-selective action of multiple photoreceptors for photomovement of the thermophilic cyanobacterium *Synechococcus elongatus*. *Photochemistry and Photobiology*, **73**, 90-95.
- Kosman, D. J.** (2003) Molecular mechanisms of iron uptake in fungi. *Molecular Microbiology*, **47**, 1185-1197.
- Kozioł, A. G., Borza, T., Ishida, K.-I., Keeling, P., Lee, R. W. & Durnford, D. G.** (2007) Tracing the evolution of the light-harvesting antennae in chlorophyll *a/b*-containing organisms. *Plant Physiology*, **143**, 1802-1816.
- Kroth, P. G., Chiovitti, A., Gruber, A., Martin-Jezequel, V., Mock, T., Parker, M. S., Stanley, M. S., Kaplan, A., Caron, L., Weber, T., Maheswari, U., Armbrust, E. V. & Bowler, C.** (2008) A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS One*, **3**, 14.
- Kustka, A. B., Allen, A. E. & Morel, F. M. M.** (2007) Sequence analysis and transcriptional regulation of iron acquisition genes in two marine diatoms. *Journal of Phycology*, **43**, 715-729.
- La Fontaine, S., Quinn, J. M., Nakamoto, S. S., Page, M. D., Gohre, V., Mosely, J. L., Kropat, J. & Merchant, S.** (2002) Copper-dependent iron assimilation pathway in the model photosynthetic eukaryote *Chlamydomonas reinhardtii*. *Eukaryotic Cell*, **1**, 736-757.
- Latasa, M., Scharek, R., Le Gall, F. & Guillou, L.** (2004) Pigment suites and taxonomic groups in Prasinophyceae. *Journal of Phycology*, **40**, 1149-1155.
- Le Gall, F., Rigaut-Jalabert, F., Marie, D., Garczarek, L., Viprey, M., Gobet, A. & Vaulot, D.** (2008) Picoplankton diversity in the South-East Pacific Ocean from cultures. *Biogeosciences*, **5**, 203-214.

**Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G. W., Prosser, J. I., Schuster, S. C. & Schleper, C.** (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature*, **442**, 806-809.

**Lepère, C., Boucher, D., Jardillier, L., Domaizon, I. & Debroas, D.** (2006) Succession and regulation factors of small eukaryote community composition in a lacustrine ecosystem (Lake Pavin). *Applied and Environmental Microbiology*, **72**, 2971-2981.

**Lepère, C., Vaultot, D. & Scanlan, D. J.** (2009) Photosynthetic picoeukaryote community structure in the South East Pacific Ocean encompassing the most oligotrophic waters on Earth. *Environmental Microbiology*, **11**, 3105-3117.

**Lepère, C., Demura, M., Kawachi, M., Romac, S., Probert, I. & Vaultot, D.** (2011) Whole-genome amplification (WGA) of marine photosynthetic eukaryote populations. *FEMS Microbiology Ecology*, **76**, 513-523.

**Lewin, J. & Reimann, B. E. F.** (1969) Silicon and plant growth. *Annual Review of Plant Physiology*, **20**, 289-304.

**Li, W.** (2009) Analysis and comparison of very large metagenomes with fast clustering and functional annotation. *BMC Bioinformatics*, **10**.

**Li, W. & Godzik, A.** (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658-1659.

**Li, W. K. W.** (1994) Primary production of prochlorophytes, cyanobacteria, and eukaryotic ultraphytoplankton - Measurements from flow cytometric sorting. *Limnology and Oceanography*, **39**, 169-175.

**Lim, E. L., Caron, D. A. & Delong, E. F.** (1996) Development and field application of a quantitative method for examining natural assemblages of protists with oligonucleotide probes. *Applied and Environmental Microbiology*, **62**, 1416-1423.

**Liu, H., Probert, I., Uitz, J., Claustre, H., Aris-Brosou, S., Frada, M., Not, F. & de Vargas, C.** (2009) Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 12803-12808.

**Longhurst, A.** (1995) Seasonal cycles of pelagic production and consumption. *Progress in Oceanography*, **36**, 77-167.

**Longhurst, A.** (2007) *Ecological Geography of the Sea*. Second Edition. Academic Press, San Diego, 542 pp.

**Lopez-Garcia, P., Philippe, H., Gail, F. & Moreira, D.** (2003) Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 697-702.

**Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A. W., Gross, O., Grumann, S., Hermann, S., Jost, R., Konig, A., Liss, T., Lussmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A. & Schleifer, K. H.** (2004) ARB: a software environment for sequence data. *Nucleic Acids Research*, **32**, 1363-1371.

**Mackey, M. D., Mackey, D. J., Higgins, H. W. & Wright, S. W.** (1996) CHEMTAX - A program for estimating class abundances from chemical markers: Application to HPLC measurements of phytoplankton. *Marine Ecology-Progress Series*, **144**, 265-283.

**Magurran, A.E.** (2003) *Measuring Biological Diversity*. Wiley-Blackwell. 260pp

**Man-Aharonovich, D., Philosof, A., Kirkup, B. C., Le Gall, F., Yogev, T., Berman-Frank, I., Polz, M. F., Vaultot, D. & Beja, O.** (2010) Diversity of active marine picoeukaryotes in the Eastern Mediterranean Sea unveiled using photosystem-II *psbA* transcripts. *ISME Journal*, **4**, 1044-1052.

**Marie, D., Partensky, F., Jacquet, S. & Vaultot, D.** (1997) Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. *Applied and Environmental Microbiology*, **63**, 186-193.

**Marie, D., Shi, X. L., Rigaut-Jalabert, F. & Vaultot, D.** (2010) Use of flow cytometric sorting to better assess the diversity of small photosynthetic eukaryotes in the English Channel. *FEMS Microbiology Ecology*, **72**, 165-178.

**Marin, B. & Melkonian, M.** (2010) Molecular phylogeny and classification of the Mamiellophyceae class. nov (Chlorophyta) based on sequence comparisons of the nuclear- and plastid-encoded rRNA operons. *Protist*, **161**, 304-336.

**Martin, J.H.** (1990) Glacial-Interglacial CO<sub>2</sub> change: The iron hypothesis. *Paleoceanography*, **5**, 1-13

**Masquelier, S. & Vaultot, D.** (2008) Distribution of micro-organisms along a transect in the South-East Pacific Ocean (BIOSOPÉ cruise) using epifluorescence microscopy. *Biogeosciences*, **5**, 311-321.

**Massana, R. & Pedros-Alio, C.** (2008) Unveiling new microbial eukaryotes in the



surface ocean. *Current Opinion in Microbiology*, **11**, 213-218.

**Massana, R., Balague, V., Guillou, L. & Pedros-Alio, C.** (2004) Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. *FEMS Microbiology Ecology*, **50**, 231-243.

**Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, S. Y., Mori, T., Nishida, K., Yagisawa, F., Yoshida, Y., Nishimura, Y., Nakao, S., Kobayashi, T., Momoyama, Y., Higashiyama, T., Minoda, A., Sano, M., Nomoto, H., Oishi, K., Hayashi, H., Ohta, F., Nishizaka, S., Haga, S., Miura, S., Morishita, T., Kabeya, Y., Terasawa, K., Suzuki, Y., Ishii, Y., Asakawa, S., Takano, H., Ohta, N., Kuroiwa, H., Tanaka, K., Shimizu, N., Sugano, S., Sato, N., Nozaki, H., Ogasawara, N., Kohara, Y. & Kuroiwa, T.** (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature*, **428**, 653-657.

**McCarren, J., Becker, J. W., Repeta, D. J., Shi, Y. M., Young, C. R., Malmstrom, R. R., Chisholm, S. W. & DeLong, E. F.** (2010) Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 16420-16427.

**McDonald, S. M., Sarno, D., Scanlan, D. J. & Zingone, A.** (2007) Genetic diversity of eukaryotic ultraphytoplankton in the Gulf of Naples during an annual cycle. *Aquatic Microbial Ecology*, **50**, 75-89.

**Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-Laylin, L. K., Marechal-Drouard, L., Marshall, W. F., Qu, L.-H., Nelson, D. R., Sanderfoot, A. A., Spalding, M. H., Kapitonov, V. V., Ren, Q., Ferris, P., Lindquist, E., Shapiro, H., Lucas, S. M., Grimwood, J., Schmutz, J., Cardol, P., Cerutti, H., Chanfreau, G., Chen, C.-L., Cognat, V., Croft, M. T., Dent, R., Dutcher, S., Fernandez, E., Fukuzawa, H., Gonzalez-Ballester, D., Gonzalez-Halphen, D., Hallmann, A., Hanikenne, M., Hippler, M., Inwood, W., Jabbari, K., Kalanon, M., Kuras, R., Lefebvre, P. A., Lemaire, S. D., Lobanov, A. V., Lohr, M., Manuell, A., Meier, I., Mets, L., Mittag, M., Mittelmeier, T., Moroney, J. V., Moseley, J., Napoli, C., Nedelcu, A. M., Niyogi, K., Novoselov, S. V., Paulsen, I. T., Pazour, G., Purton, S., Ral, J.-P., Riano-Pachon, D. M., Riekhof, W., Rymarquis, L., Schroda, M., Stern, D., Umen, J., Willows, R., Wilson, N., Zimmer, S. L., Allmer, J., Balk, J., Bisova, K., Chen, C.-J., Elias, M., Gendler, K., Hauser, C., Lamb, M. R., Ledford, H., Long, J. C., Minagawa, J., Page, M. D., Pan, J., Pootakham, W., Roje, S., Rose, A., Stahlberg, E., Terauchi, A. M., Yang, P., Ball, S., Bowler, C., Dieckmann, C. L., Gladyshev, V. N., Green, P., Jorgensen, R., Mayfield, S., Mueller-Roeber, B., Rajamani, S., Sayre, R. T., Brokstein, P., Dubchak, I., Goodstein, D., Hornick, L., Huang, Y. W., Jhaveri, J., Luo, Y., Martinez, D., Ngau, W. C. A., Otilar, B., Poliakov, A., Porter, A., Szajkowski, L., Werner, G., Zhou, K., Grigoriev, I. V., Rokhsar, D. S.,**

- Grossman, A. R., Chlamydomonas, A. & Team, J. G. I. A.** (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science*, **318**, 245-251.
- Middelboe, M. & Lyck, P. G.** (2002) Regeneration of dissolved organic matter by viral lysis in marine microbial communities. *Aquatic Microbial Ecology*, **27**, 187-194.
- Mills, M. M., Moore, C. M., Langlois, R., Milne, A., Achterberg, E., Nachtigall, K., Lochte, K., Geider, R. J. & La Roche, J.** (2008) Nitrogen and phosphorus co-limitation of bacterial productivity and growth in the oligotrophic subtropical North Atlantic. *Limnology and Oceanography*, **53**, 824-834.
- Mitra, M., Lato, S. M., Ynalvez, R. A., Xiao, Y. & Moroney, J. V.** (2004) Identification of a new chloroplast carbonic anhydrase in *Chlamydomonas reinhardtii*. *Plant Physiology*, **135**, 173-182.
- Mock, T., Krell, A., Glockner, G., Kolukisaoglu, U. & Valentin, K.** (2006) Analysis of expressed sequence tags (ESTs) from the polar diatom *Fragilariopsis cylindrus*. *Journal of Phycology*, **42**, 78-85.
- Mock, T., Samanta, M. P., Iverson, V., Berthiaume, C., Robison, M., Holtermann, K., Durkin, C., BonDurant, S. S., Richmond, K., Rodesch, M., Kallas, T., Huttlin, E. L., Cerrina, F., Sussmann, M. R. & Armbrust, E. V.** (2008) Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioprocesses. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 1579-1584.
- Montsant, A., Allen, A. E., Coesel, S., De Martino, A., Falciatore, A., Mangogna, M., Siaut, M., Heijde, M., Jabbari, K., Maheswari, U., Rayko, E., Vardi, A., Apt, K. E., Berges, J. A., Chiovitti, A., Davis, A. K., Thamatrakoln, K., Hadi, M. Z., Lane, T. W., Lippmeier, J. C., Martinez, D., Parker, M. S., Pazour, G. J., Saito, M. A., Rokhsar, D. S., Armbrust, E. V. & Bowler, C.** (2007) Identification and comparative genomic analysis of signalling and regulatory components in the diatom *Thalassiosira pseudonana*. *Journal of Phycology*, **43**, 585-604.
- Moon-van der Staay, S. Y., De Wachter, R. & Vault, D.** (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature*, **409**, 607-610.
- Moon-van der Staay, S. Y., van der Staay, G. W. M., Guillou, L., Vault, D., Claustre, H. & Medlin, L. K.** (2000) Abundance and diversity of prymnesiophytes in the picoplankton community from the equatorial Pacific Ocean inferred from 18S rDNA sequences. *Limnology and Oceanography*, **45**, 98-109.

- Moore, C. M., Mills, M. M., Langlois, R., Milne, A., Achterberg, E. P., La Roche, J. & Geider, R. J.** (2008) Relative influence of nitrogen and phosphorus availability on phytoplankton physiology and productivity in the oligotrophic sub-tropical North Atlantic Ocean. *Limnology and Oceanography*, **53**, 291-305.
- Moore, J. K., Doney, S. C., Glover, D. M. & Fung, I. Y.** (2002) Iron cycling and nutrient-limitation patterns in surface waters of the World Ocean. *Deep-Sea Research Part II-Topical Studies in Oceanography*, **49**, 463-507.
- Moreira, D. & Lopez-Garcia, P.** (2002) The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends in Microbiology*, **10**, 31-38.
- Moreira, D. & Philippe, H.** (2001) Sure facts and open questions about the origin and evolution of photosynthetic plastids. *Research in Microbiology*, **152**, 771-780.
- Moroney, J. V. & Somanchi, A.** (1999) How do algae concentrate CO<sub>2</sub> to increase the efficiency of photosynthetic carbon fixation? *Plant Physiology*, **119**, 9-16.
- Murphy, L. S. & Haugen, E. M.** (1985) The distribution and abundance of phototrophic ultraplankton in the North Atlantic. *Limnology and Oceanography*, **30**, 47-58.
- Nodwell, L. M. & Price, N. M.** (2001) Direct use of inorganic colloidal iron by marine mixotrophic phytoplankton. *Limnology and Oceanography*, **46**, 765-777.
- Noguchi, H., Park, J. & Takagi, T.** (2006) MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Research*, **34**, 5623-5630.
- Not, F., Latasa, M., Marie, D., Cariou, T., Vaultot, D. & Simon, N.** (2004) A single species, *Micromonas pusilla* (Prasinophyceae), dominates the eukaryotic picoplankton in the western English channel. *Applied and Environmental Microbiology*, **70**, 4064-4072.
- Not, F., Zapata, M., Pazos, Y., Campana, E., Doval, M. & Rodriguez, F.** (2007) Size-fractionated phytoplankton diversity in the NW Iberian coast: a combination of microscopic, pigment and molecular analyses. *Aquatic Microbial Ecology*, **49**, 255-265.
- Not, F., Latasa, M., Scharek, R., Viprey, M., Karleskind, P., Balague, V., Ontoria-Oviedo, I., Cumino, A., Goetze, E., Vaultot, D. & Massana, R.** (2008) Protistan assemblages across the Indian Ocean, with a specific emphasis on the picoeukaryotes. *Deep-Sea Research Part I-Oceanographic Research Papers*, **55**, 1456-1473.

- Palatinszky, M., Nikolausz, M., Svab, D. & Marialigeti, K.** (2011) Preferential ligation during TA-cloning of multitemplate PCR products - A factor causing bias in microbial community structure analysis. *Journal of Microbiological Methods*, **85**, 131-136.
- Palenik, B., Brahamsha, B., Larimer, F. W., Land, M., Hauser, L., Chain, P., Lamerdin, J., Regala, W., Allen, E. E., McCarren, J., Paulsen, I., Dufresne, A., Partensky, F., Webb, E. A. & Waterbury, J.** (2003) The genome of a motile marine *Synechococcus*. *Nature*, **424**, 1037-1042.
- Palenik, B., Ren, Q., Dupont, C. L., Myers, G. S., Heidelberg, J. F., Badger, J. H., Madupu, R., Nelson, W. C., Brinkac, L. M., Dodson, R. J., Durkin, A. S., Daugherty, S. C., Sullivan, S. A., Khouri, H., Mohamoud, Y., Halpin, R. & Paulsen, I. T.** (2006) Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 13555-13559.
- Palenik, B., Grimwood, J., Aerts, A., Rouze, P., Salamov, A., Putnam, N., Dupont, C., Jorgensen, R., Derelle, E., Rombauts, S., Zhou, K. M., Otilar, R., Merchant, S. S., Podell, S., Gaasterland, T., Napoli, C., Gendler, K., Manuell, A., Tai, V., Vallon, O., Piganeau, G., Jancek, S., Heijde, M., Jabbari, K., Bowler, C., Lohr, M., Robbins, S., Werner, G., Dubchak, I., Pazour, G. J., Ren, Q. H., Paulsen, I., Delwiche, C., Schmutz, J., Rokhsar, D., Van de Peer, Y., Moreau, H. & Grigoriev, I. V.** (2007) The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 7705-7710.
- Parker, M. S., Mock, T. & Armbrust, E. V.** (2008) Genomic insights into marine microalgae. *Annual Review of Genetics*, **42**, 619-645.
- Peers, G. & Niyogi, K. K.** (2008) Pond scum genomics: The genomes of *Chlamydomonas* and *Ostreococcus*. *Plant Cell*, **20**, 502-507.
- Peers, G. & Price, N. M.** (2006) Copper-containing plastocyanin used for electron transport by an oceanic diatom. *Nature*, **441**, 341-344.
- Perez, V., Fernandez, E., Maranon, E., Moran, X. A. G. & Zubkov, M. V.** (2006) Vertical distribution of phytoplankton biomass, production and growth in the Atlantic subtropical gyres. *Deep-Sea Research Part I-Oceanographic Research Papers*, **53**, 1616-1634.
- Phillips, N., Calhoun, S., Moustafa, A., Bhattacharya, D. & Braun, E. L.** (2008) Genomic insights into evolutionary relationships among heterokont lineages emphasizing the Phaeophyceae. *Journal of Phycology*, **44**, 15-18.

- Piganeau, G. & Moreau, H.** (2007) Screening the Sargasso Sea metagenome for data to investigate genome evolution in *Ostreococcus* (Prasinophyceae, Chlorophyta). *Gene*, **406**, 184-190.
- Piganeau, G., Desdevises, Y., Derelle, E. & Moreau, H.** (2008) Picoeukaryotic sequences in the Sargasso Sea metagenome. *Genome Biology*, **9**.
- Platt, T. & Sathyendranath, S.** (1988) Oceanic primary production - estimation by remote sensing at local and regional scales. *Science*, **241**, 1613-1620.
- Platt, T. & Sathyendranath, S.** (1999) Spatial structure of pelagic ecosystem processes in the global ocean. *Ecosystems*, **2**, 384-394.
- Pommier, T., Canback, B., Riemann, L., Bostrom, K. H., Simu, K., Lundberg, P., Tunlid, A. & Hagstrom, A.** (2007) Global patterns of diversity and community structure in marine bacterioplankton. *Molecular Ecology*, **16**, 867-880.
- Poretsky, R. S., Bano, N., Buchan, A., LeClerc, G., Kleikemper, J., Pickering, M., Pate, W. M., Moran, M. A. & Hollibaugh, J. T.** (2005) Analysis of microbial gene transcripts in environmental samples. *Applied and Environmental Microbiology*, **71**, 4121-4126.
- Poretsky, R. S., Hewson, I., Sun, S., Allen, A. E., Zehr, J. P. & Moran, M. A.** (2009) Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environmental Microbiology*, **11**, 1358-1375.
- Post, W. M., Peng, T. H., Emanuel, W. R., King, A. W., Dale, V. H. & Deangelis, D. L.** (1990) The global carbon cycle. *American Scientist*, **78**, 310-326.
- Potter, D., Lajeunesse, T. C., Saunders, G. W. & Anderson, R. A.** (1997) Convergent evolution masks extensive biodiversity among marine coccoid picoplankton. *Biodiversity and Conservation*, **6**, 99-107.
- Poulton, A. J., Holligan, P. M., Hickman, A., Kim, Y.-N., Adey, T. R., Stinchcombe, M. C., Holeton, C., Root, S. & Woodward, E. M. S.** (2006) Phytoplankton carbon fixation, chlorophyll-biomass and diagnostic pigments in the Atlantic Ocean. *Deep-Sea Research Part II-Topical Studies in Oceanography*, **53**, 1593-1610.
- Prochnik, S. E., Umen, J., Nedelcu, A. M., Hallmann, A., Miller, S. M., Nishii, I., Ferris, P., Kuo, A., Mitros, T., Fritz-Laylin, L. K., Hellsten, U., Chapman, J., Simakov, O., Rensing, S. A., Terry, A., Pangilinan, J., Kapitonov, V., Jurka, J., Salamov, A., Shapiro, H., Schmutz, J., Grimwood, J., Lindquist, E., Lucas, S., Grigoriev, I. V., Schmitt, R., Kirk, D. & Rokhsar, D. S.** (2010) Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science*, **329**, 223-226.

- Ptacinik, R., Solimini, A. G., Andersen, T., Tamminen, T., Brettum, P., Lepisto, L., Willen, E. & Rekolainen, S.** (2008) Diversity predicts stability and resource use efficiency in natural phytoplankton communities. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 5134-5138.
- Ragni, M. & D'Alcala, M. R.** (2004) Light as an information carrier underwater. *Journal of Plankton Research*, **26**, 433-443.
- Ram, R. J., VerBerkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., Blake, R. C., Shah, M., Hettich, R. L. & Banfield, J. F.** (2005) Community proteomics of a natural microbial biofilm. *Science*, **308**, 1915-1920.
- Rappe, M. S., Suzuki, M. T., Vergin, K. L. & Giovannoni, S. J.** (1998) Phylogenetic diversity of ultraplankton plastid small-subunit rRNA genes recovered in environmental nucleic acid samples from the Pacific and Atlantic coasts of the United States. *Applied and Environmental Microbiology*, **64**, 294-303.
- Raven, J. A., Cockell, C. S. & De La Rocha, C. L.** (2008) The evolution of inorganic carbon concentrating mechanisms in photosynthesis. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **363**, 2641-2650.
- Reinfelder, J. R.** (2011) Carbon Concentrating Mechanisms in eukaryotic marine phytoplankton. In Carlson, C. A.. (Ed.) *Annual Review of Marine Science*, Vol 3.
- Reinfelder, J. R., Milligan, A. J. & Morel, F. M. M.** (2004) The role of the C-4 pathway in carbon accumulation and fixation in a marine diatom. *Plant Physiology*, **135**, 2106-2111.
- Reyes-Prieto, A., Weber, A. P. M. & Bhattacharya, D.** (2007) The origin and establishment of the plastid in algae and plants. *Annual Review of Genetics*, **41**, 147-168.
- Riesenfeld, C. S., Schloss, P. D. & Handelsman, J.** (2004) Metagenomics: Genomic analysis of microbial communities. *Annual Review of Genetics*, **38**, 525-552.
- Robinson, C., Holligan, P., Jickells, T. & Lavender, S.** (2009) The Atlantic Meridional Transect Programme (1995-2012) Foreword. *Deep-Sea Research Part II-Topical Studies in Oceanography*, **56**, 895-898.
- Rocap, G., Distel, D. L., Waterbury, J. B. & Chisholm, S. W.** (2002) Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Applied and Environmental Microbiology*, **68**, 1180-1191.

**Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., Arellano, A., Coleman, M., Hauser, L., Hess, W. R., Johnson, Z. I., Land, M., Lindell, D., Post, A. F., Regala, W., Shah, M., Shaw, S. L., Steglich, C., Sullivan, M. B., Ting, C. S., Tolonen, A., Webb, E. A., Zinser, E. R. & Chisholm, S. W. (2003)** Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature*, **424**, 1042-1047.

**Rodriguez-Ezpeleta, N., Brinkmann, H., Burey, S. C., Roure, B., Burger, G., Löffelhardt, W., Bohnert, H. J., Philippe, H. & Lang, B. F. (2005)** Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. *Current Biology*, **15**, 1325-1330.

**Romari, K. & Vaultot, D. (2004)** Composition and temporal variability of picoeukaryote communities at a coastal site of the English Channel from 18S rDNA sequences. *Limnology and Oceanography*, **49**, 784-798.

**Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yoosheph, S., Wu, D., Eisen, J. A., Hoffman, J. M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, H., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J. E., Li, K., Kravitz, S., Heidelberg, J. F., Utterback, T., Rogers, Y.-H., Falcon, L. I., Souza, V., Bonilla-Rosso, G., Eguiarte, L. E., Karl, D. M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M. R., Strausberg, R. L., Neilson, K., Friedman, R., Frazier, M. & Venter, J. C. (2007)** The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biology*, **5**, 398-431.

**Scala, S., Carels, N., Falciatore, A., Chiusano, M. L. & Bowler, C. (2002)** Genome properties of the diatom *Phaeodactylum tricornutum*. *Plant Physiology*, **129**, 993-1002.

**Scanlan, D. J. & West, N. J. (2002)** Molecular ecology of the marine cyanobacterial genera *Prochlorococcus* and *Synechococcus*. *FEMS Microbiology Ecology*, **40**, 1-12.

**Scanlan, D. J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W. R., Post, A. F., Hagemann, M., Paulsen, I. & Partensky, F. (2009)** Ecological genomics of marine picocyanobacteria. *Microbiology and Molecular Biology Reviews*, **73**, 249-299.

**Scarsella, M., Parisi, M. P., D'Urso, A., De Filippis, P., Opoka, J. & Bravi, M. (2009)** Achievements and perspectives in hetero- and mixotrophic culturing of microalgae. In Pierucci, S. (Ed.) *Icheap-9: 9th International Conference on Chemical and Process Engineering, Pts 1-3*.

**Schmeisser, C., Steele, H. & Streit, W. R. (2007)** Metagenomics, biotechnology with non-culturable microbes. *Applied Microbiology and Biotechnology*, **75**, 955-962.

- Schneider, T. & Riedel, K.** (2010) Environmental proteomics: Analysis of structure and function of microbial communities. *Proteomics*, **10**, 785-798.
- Schnepf, E. & Elbrachter, M.** (1988) Cryptophycean like double membrane bound chloroplasts in the Dinoflagellate, *Dinophysis ehrenb* - Evolutionary, phylogenetic and toxicological implications. *Botanica Acta*, **101**, 196-203.
- Schonhuber, W., Fuchs, B., Juretschko, S. & Amann, R.** (1997) Improved sensitivity of whole-cell hybridization by the combination of horseradish peroxidase-labeled oligonucleotides and tyramide signal amplification. *Applied and Environmental Microbiology*, **63**, 3268-3273.
- Seshadri, R., Kravitz, S. A., Smarr, L., Gilna, P. & Frazier, M.** (2007) CAMERA: A community resource for metagenomics. *PLoS Biology*, **5**, 394-397.
- Shalapyonok, A., Olson, R. J. & Shalapyonok, L. S.** (2001) Arabian Sea phytoplankton during Southwest and Northeast monsoons 1995: composition, size structure and biomass from individual cell properties measured by flow cytometry. *Deep-Sea Research Part II-Topical Studies in Oceanography*, **48**, 1231-1261.
- Sharrock, R. A.** (2008) The phytochrome red/far-red photoreceptor superfamily. *Genome Biology*, **9**.
- Shi, J., Pan, K., Yu, J., Zhu, B., Yang, G., Yu, W. & Zhang, X.** (2008) Analysis of expressed sequence tags from the marine microalga *Nannochloropsis oculata* (Eustigmatophyceae). *Journal of Phycology*, **44**, 99-102.
- Shi, X. L., Marie, D., Jardillier, L., Scanlan, D. J. & Vaultot, D.** (2009) Groups without cultured representatives dominate eukaryotic picophytoplankton in the oligotrophic south east Pacific Ocean. *PLoS One*, **4**.
- Shi, X. L., Lepere, C., Scanlan, D. J. & Vaultot, D.** (2011) Plastid 16S rRNA gene diversity among eukaryotic picophytoplankton sorted by flow cytometry from the South Pacific Ocean. *PLoS One*, **6**.
- Simon, C. & Daniel, R.** (2011) Metagenomic analyses: Past and future trends. *Applied and Environmental Microbiology*, **77**, 1153-1161.
- Simon, N., Lebot, N., Marie, D., Partensky, F. & Vaultot, D.** (1995) Fluorescent in-situ hybridization with ribosomal RNA targeted oligonucleotide probes to identify small phytoplankton by flow cytometry. *Applied and Environmental Microbiology*, **61**, 2506-2513.



- Slapeta, J., Lopez-Garcia, P. & Moreira, D.** (2006) Global dispersal and ancient cryptic species in the smallest marine eukaryotes. *Molecular Biology and Evolution*, **23**, 23-29.
- Smith, V. H.** (2007) Microbial diversity-productivity relationships in aquatic ecosystems. *FEMS Microbiology Ecology*, **62**, 181-186.
- Soto, A. R., Zheng, H., Shoemaker, D., Rodriguez, J., Read, B. A. & Wahlund, T. M.** (2006) Identification and preliminary characterization of two cDNAs encoding unique carbonic anhydrases from the marine alga *Emiliania huxleyi*. *Applied and Environmental Microbiology*, **72**, 5500-5511.
- Sowell, S. M., Wilhelm, L. J., Norbeck, A. D., Lipton, M. S., Nicora, C. D., Barofsky, D. F., Carlson, C. A., Smith, R. D. & Giovanonni, S. J.** (2009) Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *ISME Journal*, **3**, 93-105.
- Spalding, M. H.** (2008) Microalgal carbon-dioxide-concentrating mechanisms: *Chlamydomonas* inorganic carbon transporters. *Journal of Experimental Botany*.
- Stadtman, T. C.** (1996) Selenocysteine. *Annual Review of Biochemistry*, **65**, 83-100.
- Steglich, C., Futschik, M. E., Lindell, D., Voss, B., Chisholm, S. W. & Hess, W. R.** (2008) The challenge of regulation in a minimal photoautotroph: non-coding RNAs in *Prochlorococcus*. *PLoS Genetics*, **4**.
- Stepanauskas, R. & Sieracki, M. E.** (2007) Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 9052-9057.
- Takahashi, H., Kopriva, S., Giordano, M., Saito, K. & Hell, R.** (2011) Sulfur assimilation in photosynthetic organisms: molecular functions and regulations of transporters and assimilatory enzymes. *Annual Review of Plant Biology*, **62**, 157-184
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S.** (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*. doi:10.1093/molbev/msr121
- Tarran, G. A., Heywood, J. L. & Zubkov, M. V.** (2006) Latitudinal changes in the standing stocks of nano- and picoeukaryotic phytoplankton in the Atlantic Ocean. *Deep-Sea Research Part II-Topical Studies in Oceanography*, **53**, 1516-1529.

**Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J. & Natale, D. A.** (2003) The COG database: An updated version includes eukaryotes. *BMC Bioinformatics*, **4**.

**Teira, E., Mourino, B., Maranon, E., Perez, V., Pazo, M. J., Serret, P., de Armas, D., Escanez, J., Woodward, E. M. S. & Fernandez, E.** (2005) Variability of chlorophyll and primary production in the Eastern North Atlantic Subtropical Gyre: potential factors affecting phytoplankton activity. *Deep-Sea Research Part I-Oceanographic Research Papers*, **52**, 569-588.

**Tengs, T., Dahlberg, O. J., Shalchian-Tabrizi, K., Klaveness, D., Rudi, K., Delwiche, C. F. & Jakobsen, K. S.** (2000) Phylogenetic analyses indicate that the 19 ' hexanoyloxy-fucoanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Molecular Biology and Evolution*, **17**, 718-729.

**Terauchi, M., Kato, A., Nagasato, C. & Motomura, T.** (2010) Research note: Analysis of expressed sequence tags from the chrysophycean alga *Ochromonas danica* (Heterokontophyta). *Phycological Research*, **58**, 217-221.

**Ter Braak, C. J. F.** (1986) Canonical correspondence analysis - A new eigenvector technique for multivariate direct gradient analysis. *Ecology*, **67**, 1167-1179.

**Tirichine, L. & Bowler, C.** (2011) Decoding algal genomes: tracing back the history of photosynthetic life on Earth. *Plant Journal*, **66**, 45-57.

**Tompkins, J., DeVille, M.M., Day, J.G. & Turner, M.F.** (1995) Culture collection of algae and protozoa. Catalog of strains. Ambleside, UK, 204 pp.

**Tortell, P. D.** (2000) Evolutionary and ecological perspectives on carbon acquisition in phytoplankton. *Limnology and Oceanography*, **45**, 744-750.

**Tringe, S. G. & Rubin, E. M.** (2005) Metagenomics: DNA sequencing of environmental samples. *Nature Reviews Genetics*, **6**, 805-814.

**Tringe, S. G., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., Podar, M., Short, J. M., Mathur, E. J., Detter, J. C., Bork, P., Hugenholtz, P. & Rubin, E. M.** (2005) Comparative metagenomics of microbial communities. *Science*, **308**, 554-557.

**Tyrrell, T.** (1999) The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature*, **400**, 525-531.

**Tyson, G. W. & Banfield, J. F.** (2005) Cultivating the uncultivated: a community genomics perspective. *Trends in Microbiology*, **13**, 411-415.

**Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., Solovyev, V. V., Rubin, E. M., Rokhsar, D. S. & Banfield, J. F.** (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, **428**, 37-43.

**Urlich, T., Lanzen, A., Qi, J., Huson, D. H., Schleper, C. & Schuster, S. C.** (2008) Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One*, **3**.

**Vaulot, D., Romari, K. & Not, F.** (2002) Are autotrophs less diverse than heterotrophs in marine picoplankton? *Trends in Microbiology*, **10**, 266-267.

**Vaulot, D., Eikrem, W., Viprey, M. & Moreau, H.** (2008) The diversity of small eukaryotic phytoplankton ( $\leq 3 \mu\text{m}$ ) in marine ecosystems. *FEMS Microbiology Reviews*, **32**, 795-820.

**Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D. Y., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y. H. & Smith, H. O.** (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, **304**, 66-74.

**Viprey, M., Guillou, L., Ferreol, M. & Vaulot, D.** (2008) Wide genetic diversity of picoplanktonic green algae (Chloroplastida) in the Mediterranean Sea uncovered by a phylum-biased PCR approach. *Environmental Microbiology*, **10**, 1804-1822.

**Vogel, J., Bartels, V., Tang, T. H., Churakov, G., Slagter-Jager, J. G., Huttenhofer, A. & Wagner, E. G. H.** (2003) RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Research*, **31**, 6435-6443.

**Warnecke, F. & Hugenholtz, P.** (2007) Building on basic metagenomics with complementary technologies. *Genome Biology*, **8**, 231

**Watanabe, M. M., Suda, S., Inouye, I., Sawaguchi, T. & Chihara, M.** (1990) *Lepidodinium viride* gen. et sp. nov. (Gymnodiniales, Dinophyta), a green dinoflagellate with a chlorophyll *a* and *b* containing endosymbiont. *Journal of Phycology*, **26**, 741-751.

Watve, M., Shejval, V., Sonawane, C., Rahalkar, M., Matapurkar, A., Shouche, Y., Patole, M., Phadnis, N., Champhenkar, A., Damle, K., Karandikar, S., Kshirsagar, V. & Jog, M. (2000) The 'K' selected oligophilic bacteria: A key to uncultured diversity? *Current Science*, **78**, 1535-1542.

West, N. J., Schonhuber, W. A., Fuller, N. J., Amann, R. I., Rippka, R., Post, A. F. & Scanlan, D. J. (2001) Closely related *Prochlorococcus* genotypes show remarkably different depth distributions in two oceanic regions as revealed by in situ hybridization using 16S rRNA-targeted oligonucleotides. *Microbiology*, **147**, 1731-1744.

West, S. A., Diggle, S. P., Buckling, A., Gardner, A. & Griffins, A. S. (2007) The social lives of microbes. *Annual Review of Ecology Evolution and Systematics*, **38**, 53-77.

Worden, A. Z. & Allen, A. E. (2010) The voyage of the microbial eukaryote. *Current Opinion in Microbiology*, **13**, 652-660.

Worden, A.Z., & Not, F. (2008) Ecology and diversity of picoeukaryotes. In *Microbial Ecology of the Oceans*, Second edition, pp. 159-205 Edited by Kirchman, D.L. Wiley and Sons, Inc

Worden, A. Z., Nolan, J. K. & Palenik, B. (2004) Assessing the dynamics and ecology of marine picophytoplankton: The importance of the eukaryotic component. *Limnology and Oceanography*, **49**, 168-179.

Worden, A. Z., Lee, J. H., Mock, T., Rouze, P., Simmons, M. P., Aerts, A. L., Allen, A. E., Cuvelier, M. L., Derelle, E., Everett, M. V., Foulon, E., Grimwood, J., Gundlach, H., Henrissat, B., Napoli, C., McDonald, S. M., Parker, M. S., Rombauts, S., Salamov, A., Von Dassow, P., Badger, J. H., Coutinho, P. M., Demir, E., Dubchak, I., Gentemann, C., Eikrem, W., Gready, J. E., John, U., Lanier, W., Lindquist, E. A., Lucas, S., Mayer, K. F. X., Moreau, H., Not, F., Otilar, R., Panaud, O., Pangilinan, J., Paulsen, I., Piegu, B., Poliakov, A., Robbens, S., Schmutz, J., Toulza, E., Wyss, T., Zelensky, A., Zhou, K., Armbrust, E. V., Bhattacharya, D., Goodenough, U. W., de Peer, Y. V. & Grigoriev, I. V. (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science*, **324**, 268-272.

Wright, S. W., Ishikawa, A., Marchant, H. J., Davidson, A. T., van den Enden, R. L. & Nash, G. V. (2009) Composition and significance of picophytoplankton in Antarctic waters. *Polar Biology*, **32**, 797-808.

Xiong, J., Fischer, W. M., Inoue, K., Nakahara, M. & Bauer, C. E. (2000) Molecular evidence for the early evolution of photosynthesis. *Science*, **289**, 1724-1730.

**Yooseph, S., Sutton, G., Rusch, D. B., Halpern, A. L., Williamson, S. J., Remington, K., Eisen, J. A., Heidelberg, K. B., Manning, G., Li, W. Z., Jaroszewski, L., Cieplak, P., Miller, C. S., Li, H. Y., Mashiyama, S. T., Joachimiak, M. P., van Belle, C., Chandonia, J. M., Soergel, D. A., Zhai, Y. F., Natarajan, K., Lee, S., Raphael, B. J., Bafna, V., Friedman, R., Brenner, S. E., Godzik, A., Eisenberg, D., Dixon, J. E., Taylor, S. S., Strausberg, R. L., Frazier, M. & Venter, J. C. (2007) The Sorcerer II Global Ocean Sampling expedition: Expanding the universe of protein families. *PLoS Biology*, **5**, 432-466.**

**Zeidner, G., Preston, C. M., Delong, E. F., Massana, R., Post, A. F., Scanlan, D. J. & Beja, O. (2003) Molecular diversity among marine picophytoplankton as revealed by *psbA* analyses. *Environmental Microbiology*, **5**, 212-216.**

**Zengler, K., Toledo, G., Rappe, M., Elkins, J., Mathur, E. J., Short, J. M. & Keller, M. (2002) Cultivating the uncultured. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 15681-15686.**

**Zhu, F., Massana, R., Not, F., Marie, D. & Vaulot, D. (2005) Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiology Ecology*, **52**, 79-92.**

**Zubkov, M. V. & Tarran, G. A. (2008) High bacterivory by the smallest phytoplankton in the North Atlantic Ocean. *Nature*, **455**, 224-U48.**

**Zubkov, M., Burkill, P. H. & Topping, J. N. (2007) Flow cytometric enumeration of DNA-stained oceanic planktonic protists. *Journal of Plankton Research*, **29**, 79-86.**

**Zwirgmaier, K., Heywood, J. L., Chamberlain, K., Woodward, E. M. S., Zubkov, M. V. & Scanlan, D. J. (2007) Basin-scale distribution patterns of picocyanobacterial lineages in the Atlantic Ocean. *Environmental Microbiology*, **9**, 1278-1290.**

## Appendix